(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 November 2003 (13.11.2003)

PCT

(10) International Publication Number
WO 03/093307 A2

(51) International Patent Classification[7]: C07K 14/35, G01N 33/569, C12N 15/31, A61K 38/16, 39/04

(21) International Application Number: PCT/GB03/01815

(22) International Filing Date: 28 April 2003 (28.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0209723.6 27 April 2002 (27.04.2002) GB
0209724.4 27 April 2002 (27.04.2002) GB

(71) Applicant *(for all designated States except US)*: **THE SECRETARY OF STATE FOR ENVIRONMENT, FOOD & RURAL AFFAIRS** [GB/GB]; (DEFRA), Nobel House, 17 Smith Square, London SW1P 3JR (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **COCKLE, Paul, Jason** [GB/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB). **VORDERMEIER, Hanns, Martin** [DE/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB). **GORDON, Stephen, Vincent** [DE/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB). **HEWINSON, Robert, Glyn** [DE/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB).

(74) Agent: **GREAVES, Carol, Pauline**; Greaves Brewster, Indigo House, Cheddar Business Park, Wedmore Road, Cheddar, Somerset BS27 3EB (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/093307 A2

(54) Title: MYCOBACTERIAL ANTIGENS AND USES THEREOF

(57) Abstract: The present invention relates to the use of antigens derived from the RD1 or RD2 regions of the *Mycobacterium tuberculosis*, *Mycobacterium bovis* or *Mycobacterium africanum* genomes, and peptides derived therefrom, as diagnostic reagents, in particular in the context of diagnostic kits. In addition, certain of these peptides, as well as other antigens and peptides derived from the RD14 region of the genome are suitable for use as vaccines. Novel fusion peptides are also part of the invention.

Antigens and Uses Thereof

The present invention relates to the use of antigens derived from the RD1 or RD2 regions of the *Mycobacterium tuberculosis*, *Mycobacterium bovis* or *Mycobacterium africanum* genomes, and peptides derived therefrom, as diagnostic reagents, in particular in the context of diagnostic kits. In addition, certain of these peptides, as well as other antigens and peptides derived from the RD14 region of the genome are suitable for use as vaccines. Novel fusion peptides are also part of the invention.

In particular, the present invention relates to diagnostic kits comprising such antigens for differentiating between those mammals infected by tuberculosis, those which have been vaccinated against tuberculosis, and those mammals, which have been sensitised by environmental *Mycobacteria*.

The present invention further relates to novel *Mycobacterium tuberculosis*, *Mycobacterium bovis* and *Mycobacterium africanum* peptides derived from such antigens.

The present invention also relates to vaccines against *Mycobacterium* infections, in particular, *Mycobacterium tuberculosis*, *Mycobacterium bovis* or *Mycobacterium africanum* infections, as well as to veterinary and pharmaceutical compositions containing these and their preparations.

Bovine tuberculosis (BTB) is caused by *Mycobacterium bovis* and shares greater than 99.9% DNA identity with *M. tuberculosis*, the main cause of human tuberculosis. Moreover, BTB is a zoonotic disease and was responsible during the 1930s and 1940s for approximately 6% of the total human deaths due to TB, and more than 50% of all cervical lymphadentitis cases in children. The introduction of pasteurisation of milk in the 1930s dramatically reduced the transmission from cattle to man. However, it still

remains a small but significant cause of human morbidity and mortality especially in developing countries and is seen as one of the most important infectious diseases of both man and other animals in the world.

Mycobacterium bovis causes disease in both cattle and man. In the absence of a BTB control programme, TB in cattle can have severe implications for animal welfare, causing reduced productivity and premature death, resulting in substantial economic losses to affected farms. A compulsory eradication programme based upon the slaughter of infected animals, detected by the single intradermal comparative tuberculin skin test, began in Great Britain (GB) in 1950 and by 1960 it had been implemented in all of Britain. These measures resulted in the dramatic reduction of bovine tuberculosis in GB from incidence rates of around 40% of cattle infected with *M. bovis* to 0.41% in 1996. However, despite continued implementation of these control measures, the incidence of BTB in cattle has been steadily rising since 1988, possibly due to a wildlife reservoir of *M. bovis*.

BCG is an attenuated strain of *M. bovis*, and is currently the only available vaccine for the prevention of BTB. Encouraging results with BCG have been reported in New Zealand where a significant level of protection in BCG vaccinated cattle against experimental *M. bovis* infection has been recently demonstrated.

Immunity to *M. tuberculosis* is characterised by three basic features: 1) living bacilli which efficiently induce a protective immune response; 2) specifically sensitised T lymphocytes which mediate this protection, and 3) interferon gamma (IFN-γ) which is an important mediator molecule.

Cattle with a mycobacterial infection will predominantly mount a cellular immune response. Therefore, the skin test using tuberculin PPD has become an integral part of the bovine

tuberculosis eradication programme. In addition to skin tests, blood-based diagnostic assays that measure antigen-induced lymphokine production such as the IFN-γ are also under consideration. The cytokine IFN-γ appears to be critical in the development of immunity to *M. tuberculosis.* For example, both mice with a disruptive IFN-γ gene and humans with mutated IFN-γ receptor are highly susceptible to mycobacterial infections. However, specificity constraints are associated with the use of PPD in such assays. These arise due to the crude mixture of *M. bovis* proteins that it contains, many that are cross-reactive with other environmental mycobacterial species, e.g., *M. avium or M. intracellulare* and importantly the vaccine strain *M. bovis* Bacille Calmette-Guerin (BCG).

A cattle vaccine would reduce the risk of cattle infection and hence result in lower tuberculin test frequencies and significant cost savings. It is believed that the development of an improved cattle vaccine holds the best long-term prospect for BTB control in British herds. In addition, it would be desirable to develop a complimentary diagnostic test to differentiate between vaccinated animals and those infected with M. bovis (*differential diagnosis*) in parallel with the vaccine to ensure continuation of the test and slaughter-based control strategies.

As previous studies have demonstrated, diagnostic reagents which distinguish between vaccinated and infected cattle can be developed using specific, defined antigens that are present in virulent *M. bovis* but absent from the vaccine strain. Genetic analysis of BCG has revealed that several large genomic regions have been deleted during attenuation and subsequent prolonged propagation in culture [Behr, M. A., et al.. 1999. Science 284:1520 - 1523; Gordon, S. V., et al. 1999. Mol. Microbiol. 32: 643-655]. These regions have been characterised and antigens from one of these regions, RD1 [Mahairas, G. G., et al. 1996. J. Bacteriol. 178:1274-82], have been studied extensively

in several species including humans and cattle. For example, it has been recently demonstrated that protein or peptide cocktails composed of two RD1 region antigens, ESAT-6 and CFP-10, can be used to distinguish between BCG vaccinated and M. bovis infected cattle [Van pinxteren, et al.. 2000. Clin. Diagn. Lab. Immunol. 7:155-160; Vordermeier, H. M. et al. 2001. Clin. Diagn. Lab. Immunol. 8:571-8].

However, the level of sensitivity achieved with these antigens has not reached that of tuberculin. It would, therefore, be desirable to provide other antigens in order to achieve this desired sensitivity. Such antigens may also be useful in supplementing the ESAT-6 and CFP-10 to achieve even greater sensitivity.

In alternative approach to using recombinant proteins is the application of overlapping synthetic peptides derived from those antigens described above. Synthetic peptides have the advantages of lower production costs, easier standardisation, improved quality control and carry no risk of infection since they are chemically synthesised.

Such synthetic peptide epitopes have been found to have great potential in the study of immune responses in cattle and in the development of diagnostic reagents. For example, formulation of 10 synthetic peptides derived from ESAT-6 and CFP-10 resulted in similar cellular immune responses to those seen with the whole recombinant antigens. When assayed in cattle this cocktail could distinguish between *M. bovis* infected animals and BCG vaccinated cattle with sensitivity similar to PPD and with a greater specificity [Vordermeier, 2001 supra.].

Differential diagnosis is not the only concern associated with BCG. BCG vaccination studies have highlighted the variability with regard to its efficacy. In humans, this ranges from 0 to 80% when tested in different populations, with consistently poor

results observed in the equatorial regions. Similar variations in efficacy have also been reported in BCG vaccination experiments and trials in cattle (e.g. [Buddle, B., et al, 2002. Vaccine 20: 1126-33). It would therefore be desirable to improve or supplement BCG vaccination. Strategies to generate novel tuberculosis vaccines include sub-unit vaccination with either DNA vaccines or protein subunits (Rev. [Anderson, P. 2001, TB Vaccines: progress and problems. Trends Immunol]). Antigens such as MPT-64 and ESAT-6, whose genes were deleted in BCG, have been tested as DNA vaccines and imparted protective immunity in small animal models.

The present invention seeks to provide an improved diagnostic test to differentiate between vaccinated mammals and those infected with tuberculosis. Preferably, the test of the present invention differentiates between animals vaccinated against *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum* and those infected with *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum*.

The present invention also seeks to provide an improved vaccine for control of tuberculosis and in particular to control tuberculosis in cattle. The tuberculosis disease also affects a number of other different animal species such as guinea pigs, badgers, possums and deer. The vaccines of the present invention may therefore be useful in the control of tuberculosis infections in such different animals.

The applicants have found that certain polypeptides derived from the RD1 or RD2 regions of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum* genomes, or a variant, homologue or fragment of these, other than ESAT-6, CFP-10, MPT-64 are useful as diagnostic agents, and in particular are a source of diagnostic peptides. Suitably the polypeptide is other than polypeptides encoded by the Rv1984c, Rv3871, Rv3872 or at least certain parts of the Rv3873 regions of the

Mycobacterium tuberculosis, Mycobacterium bovis or *Mycobacterium africanum* genomes.

The term "polypeptide" as used herein includes long chain peptides including proteins and epitopic fragments thereof. Such proteins generally comprise one or more chains of amino acids joined covalently through peptide bonds and are typically greater than 10,000 MW.

The term "peptides" refers to small proteins (generally less than about 10,000 MW), and in particular to smaller chains, for example up to 30 amino acids in length, preferably up to 20 amino acids in length. Also included however are small oligopeptides comprising three or more amino acid residues covalently linked through peptide bonds. Peptides will generally comprise two or more amino acid residues linked together covalently through peptide bonds.

The polypeptide from which the diagnostic agents are selected are preferably encoded by the *Mycobacterium tuberculosis* genome and comprise a member of the PE/PPE protein family.

The term "derived from" as used herein means any polypeptide or peptide encoded by an open reading frame from the specified regions of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum* genomes.

In particular, the applicants utilise peptides encoded by fragments of the open reading frames and variants thereof as long as such peptides are still capable of being used as diagnostic reagents. In particular, they will comprise epitopic sequences.

According to the present invention there is provided a diagnostic reagent comprising a peptide comprising an epitope from at least one polypeptide selected from Rv1986 (SEQ ID NO

1), Rv3878 (SEQ ID NO 3), Rv 1983 (SEQ ID NO 4), Rv3873 (SEQ ID NO 5) or Rv3879 (SEQ ID NO 6).

An "epitope" of the sequence comprises those amino acids that are necessary to generate in immune response, and therefore be recognised in a diagnostic test. They may be consecutive amino acids, or they may be spaced apart from one another. In the latter case, the nature of the amino acids between the amino acids of the epitope may not be crucial to the activity and may be varied. Determination of the amino acids which comprise the epitopes can be determined using routine methods, for example by finding antigenic regions or fragments, as illustrated hereinafter, and then carrying out a series of small mutations, for example, point mutations, and then determining whether the immunogenic or diagnostic activity has been retained. Where it has, then the variant retains the epitope. If activity has been lost, then the mutation has disrupted the epitope and so must be reversed.

Suitably, the peptide comprises a series of consecutive amino acids from within the polypeptide sequence. In one embodiment, the polypeptide from which the peptide is derived comprises the sequence shown in SEQ ID Nos 1, 3, 4 or 6.

Alternatively, the polypeptide from which the peptide is derived comprises the sequence shown in SEQ ID Nos 3, 5 or 6.

Particular examples of diagnostic reagents comprise peptides which include an epitope from SEQ ID NO 23 as shown in Figure 9 hereinafter, which is a fragment of SEQ ID NO 5, and in particular, an epitope found within SEQ ID NO 25 or SEQ ID NO 7, or within SEQ ID NOs 28 and 29. SEQ ID NO 23, and fragments thereof, such as SEQ ID NOS 7,25, 28 and 29 form particular embodiments of the diagnostic reagents of the invention.

Other particular examples of diagnostic reagents comprise peptides which include an epitope from SEQ ID NO 35, shown in Figure 10 hereinafter, which is a fragment of SEQ ID NO 3. SEQ ID NO 35 or fragments thereof, form a particular embodiment of the invention.

Particular examples of diagnostic reagents comprise peptides which include an epitope from SEQ ID NO 48, shown in Figure 11 hereinafter, which is a fragment of SEQ ID NO 6, and in particular, an epitope found within SEQ ID NO 51. SEQ ID NO 48 or variants thereof, or fragments of these, in particular SEQ ID NO 51 form particular embodiments of the diagnostic reagents of the invention.

The polypeptides themselves, as well as homologues or variants thereof may also be used as diagnostic agents, but preferably fragments (comprising peptides are employed).

The peptides described above may be used in either specific or differential diagnostic tests.

The term "fragment thereof" as used herein in relation to an amino acid sequence refers to any portion of the given amino acid sequence which has the same activity as the complete amino acid sequence. Fragments will suitably comprise at least 10 and preferably at least 20 consecutive amino acids from the basic sequence. In one embodiment, the fragment sequence comprises 17 amino acids.

Fragments of the polypeptide include deletion mutants and polypeptides where small regions of the polypeptides are joined together. The fragments should contain an epitope, and preferably contain at least one antigenic region.

The term "homologue" refers to similar genes found in other organisms, such as the *Mycobacterium bovis* or *Mycobacterium africanum* genomes,

The term "variant thereof" as used herein in relation to an amino acid sequence means sequences of amino acids which differ from the base sequence from which they are derived in that one or more amino acids within the sequence are substituted for other amino acids. Amino acid substitutions may be regarded as "conservative" where an amino acid is replaced with a different amino acid with broadly similar properties. Non-conservative substitutions are where amino acids are replaced with amino acids of a different type.

By "conservative substitution" is meant the substitution of an amino acid by another one of the same class; the classes being as follows:

CLASS	EXAMPLES OF AMINO ACID
Nonpolar:	Ala, Val, Leu, Ile, Pro, Met, Phe, Trp
Uncharged polar:	Gly, Ser, Thr, Cys, Tyr, Asn, Gln
Acidic:	Asp, Glu
Basic:	Lys, Arg, His

As is well known to those skilled in the art, altering the primary structure of a peptide by a conservative substitution may not significantly alter the activity of that peptide because the side-chain of the amino acid which is inserted into the sequence may be able to form similar bonds and contacts as the side chain of the amino acid which has been substituted out. This is so even when the substitution is in a region, which is critical in determining the peptides conformation.

Non-conservative substitutions are possible provided that these do not interrupt with the antigenicity of the polypeptide.

Broadly speaking, fewer non-conservative substitutions will be possible without altering the biological activity of the polypeptide. Suitably, variants will be at least 50% identical, 60% identical, preferably at least 75% identical, and more preferably at least 90% identical to the base sequence.

Identity in this instance can be judged for example using the algorithm of Lipman-Pearson, with Ktuple:2, gap penalty:4, Gap Length Penalty:12, standard PAM scoring matrix (Lipman, D.J. and Pearson, W.R., Rapid and Sensitive Protein Similarity Searches, *Science*, 1985, vol. 227, 1435-1441).

The applicants have found that a diagnostic test based upon SEQ ID NOS 1 and 3 or homologues or variants thereof, can give a differential diagnostic test, which in particular, can differentiate between tuberculosis-infected and tuberculosis vaccinated mammals. Selection of a diagnostic reagent comprising or derived from these sequences can be made so that they differentiate between *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum* -infected mammals and mammals vaccinated against *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum*.

Alternatively, the applicants have found that diagnostic tests based upon SEQ ID Nos 3 or 6, or a homologue or variant thereof, can also provide differential diagnostic tests, but in this case they are able to distinguish between mammals, which are either vaccinated against or infected by tuberculosis and mammals, sensitised by environmental mycobacteria. The diagnostic reagent used in a specific diagnostic test preferably differentiates between *Mycobacterium bovis* -infected and mammals sensitised by environmental mycobacteria.

Thus the applicants have found a range of diagnostic peptides derived from an RD1 or RD2 region of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum*

genomes, or a variant, homologue or fragment of these, which are additional to those derived from the ESAT-6 or CFP-10 polypeptides.

These peptides are capable of being used as diagnostic reagents and are preferably synthetic peptides having the advantages discussed above.

One such peptide is a peptide derived from SEQ ID NO.5, which is shown in Figure 6 as SEQ ID NO 7. Fragments, homologues and variants of this peptide are also included herein. The peptide as shown in SEQ ID NO 7 may be used in a specific diagnostic test to differentiate between those mammals, which are either vaccinated against or infected by tuberculosis, and those mammals which have been sensitised by environmental mycobacteria. In particular, the peptide is especially useful in differentiating between *Mycobacterium bovis*-infected mammals, such as cattle or calves, and those animals sensitised by environmental bacteria.

Such peptides may be used as diagnostic reagents, either on their own or preferably with one or more other peptides, which may be other peptides according to the present invention, or different peptides, in order to achieve greater sensitivity and specificity of a diagnostic test. For example, protein or peptide cocktails composed of other antigens from the RD1 or RD2 regions of the *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum* genomes may be used in addition to the antigens of the present invention in order to enhance the specificity of the diagnostic test. In particular, peptide cocktails may include peptides derived from the antigens, ESAT-6 and CFP-10, as well as those described above.

According to a further aspect of the present invention there is provided a diagnostic kit comprising at least two diagnostic

reagents, at least one of which is a diagnostic reagent as described above.

In particular, kits of the invention will comprise polypeptides or peptides, at least one of which is selected from a polypeptide derived from the sequences shown as SEQ ID Nos 1, 3, 4, 6 and 7, or a fragment, homologue or variant thereof, and optionally at least one polypeptide derived from the sequences shown as SEQ ID Nos 2 and 5, and optionally one or more reagents. Such kits may be used to differentiate between tuberculosis-infected and tuberculosis-vaccinated mammals.

The polypeptide and peptide sequences described herein can provide a means for detecting the recognition of the polypeptides or peptides by the T-cell. Preferably, diagnostic reagents utilised in the diagnostic kit are able to differentiate between *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum* -infected mammals and mammals vaccinated against *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum.*

Where the kit is intended to be used to differentiate between those mammals infected by *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobactrium africanum* and those mammals which have been vaccinated against *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum,* the kit will preferably comprise the polypeptides derived from the sequences shown as SEQ ID Nos 1, 2 and 3, or a fragment, homologue or variant thereof. In particular, it will comprise diagnostic reagents comprising peptides comprising an epitope from these sequences.

Where the kit is intended to be used to differentiate between those mammals infected by *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobactrium africanum* and mammals sensitised by environmental mycobacteria, the kit will preferably comprise polypeptides or peptides derived from the sequences shown as SEQ

ID Nos 4, 5, 6, and optionally, SEQ ID NO. 7 or a fragment, homologue or variant thereof. In particular, it will comprise diagnostic reagents comprising peptides comprising an epitope from these sequences.

The diagnostic kit may also comprise one or more polypeptides or peptides from the RD1 region of the *Mycobacterium bovis*, *Mycobacterium tuberculosis* or *Mycobacterium africanum* genomes. Protein or peptide cocktails composed of such polypeptides may also be used. Especially preferred are peptide cocktails composed of the ESAT-6 and/ or the CFP-10 polypeptides. Such peptide cocktails may be used to enhance the sensitivity of the diagnostic tests of the present invention.

As would be understood, the polypeptides and peptides described above are encoded by nucleic acids. Novel nucleic acids, for example which encode novel peptides or polypeptides as described above form a further aspect of the invention, together with fragments homologues or variants thereof.

The nucleic acid may be DNA or RNA, and where it is a DNA molecule, it may comprise a cDNA or genomic DNA. These nucleic acids may themselves be useful as vaccines and such vaccines form a further aspect of the present invention.

Preferably, the nucleic acid comprises the sequence shown in SEQ ID Nos 8, 10, 11 or 13, or a variant or fragment thereof.

The term "fragment thereof" as used herein in relation to a nucleic acid or polynucleotide sequence refers to any portion of the given polynucleotide sequence which exhibits the same activity as the complete polynucleotide sequence. Fragments will suitably comprise at least 15, preferably at least 30 and more preferably at least 60 consecutive bases from the basic sequence.

The term "variant thereof" in relation to a polynucleotide or nucleic acid sequences means any substitution of, variation of, modification of, replacement of deletion of, or the addition of one or more nucleic acid(s) from or to a polynucleotide sequence providing the resultant protein sequence encoded by the polynucleotide exhibits the same properties as the protein encoded by the basic sequence. The term therefore includes allelic variants and also includes a polynucleotide which substantially hybridises to the polynucleotide sequence of the present invention. Preferably, such hybridisation occurs at, or between low and high stringency conditions. In general terms, low stringency conditions can be defined as 3 x SSC at about ambient temperature to about 55°C and high stringency condition as 0.1 x SSC at about 65°C. SSC is the name of the buffer of 0.15M NaCl. 0.015M tri-sodium citrate. 3 x SSC is three times as strong as SSC and so on.

Typically, variants have 62% or more of the nucleotides in common with the polynucleotide sequence of the present invention, more typically 65%, preferably 70%, even more preferably 80% or 85% and, especially preferred are 90%, 95%, 98% or 99% or more identity.

When comparing nucleic acid sequences for the purposes of determining the degree of identity, programs such as BESTFIT and GAP (both from Wisconsin Genetics Computer Group (GCG) software package). BESTFIT, for example, compares two sequences and produces an optimal alignment of the most similar segments. GAP enables sequences to be aligned along their whole length and fins the optimal alignment by inserting spaces in either sequence as appropriate. Suitably, in the context of the present invention when discussing identity of nucleic acid sequences, the comparison is made by alignment of the sequences along their whole length.

Generally speaking, diagnosis of infection in a host, or exposure of a host, to a mycobacterium can be carried out by
i) contacting a population of cells from the host with a polypeptide derived from an RD1 or RD2 region of the *Mycobacterium tuberculosis*, *Mycobacterium bovis* or *Mycobacterium africanum* genomes, or a variant, homologue or fragment of these, which polypeptide may be used as a diagnostic reagent, including those described above, in addition to those derived from ESAT-6, CFP-10, MPT-64; and
ii) determining whether the cells of said cell population recognise the polypeptide or fragment or variant thereof.

Thus in accordance with the invention there is provided a method of diagnosising infection in a host, or exposure of a host, to a mycobacterium, said method comprising
i) contacting a population of cells from the host with a diagnostic reagent according to the invention; and
ii) determining whether the cells of said cell population recognise the diagnostic reagent.

Suitable diagnostic reagents are as described above.

The population of cells used in the method is suitably a population of T-cells. The method preferably diagnoses infection by *Mycobacterium tuberculosis*, *Mycobacterium bovis* or *Mycobacterium africanum*.

The diagnostic reagents and peptides described above can be used, in accordance with a further aspect of the invention to produce an antibody specific to the peptide.

Polypeptides of the invention may be isolated from strains of *M. bovis*, *M. tuberculosis* or *M. africanum*. Preferably, they are prepared synthetically using conventional peptide synthesisers. Alternatively, they may be produced using recombinant DNA technology or isolated from natural sources followed by any

chemical modification, if required. In these cases, nucleic acids encoding the polypeptides are incorporated into suitable expression vectors, which are then used to transform a suitable host cell, such as a prokaryotic cell such as *E. coli*. The transformed cells are cultured and the polypeptide isolated therefrom. Vectors, cells and methods of this type form further aspects of the present invention.

A particular diagnostic kit comprising the polypeptides derived from the sequences shown as SEQ ID Nos 1 to 3 and further comprising one or more polypeptides derived from the RD1 region of *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum,* and optionally one or more reagents, for differentiating between cattle infected by *M. bovis* and cattle which have been vaccinated with BCG or with a vaccine according to the present invention.

A specific diagnostic kit comprises the polypeptides and peptides derived from the sequences shown as SEQ ID Nos 4 to 7 and further comprising one or more polypeptides derived from the RD1 region of *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum,* and optionally one or more reagents, for differentiating between cattle which have either been vaccinated against or infected by *M. bovis* and those cattle which have been sensitised by environmental mycobacteria.

A further preferred embodiment of the present invention is a vaccine comprising a peptide having the sequence shown in SEQ ID No 7.

An advantage of the present invention is that the level of sensitivity achieved in diagnostic tests with these antigens is higher than the sensitivity achieved with the antigens ESAT-6 and CFP-10. In addition, the level of specificity of the antigen of the present invention is higher than that of PPD, which is currently used. PPD has the disadvantage that it

cross-reacts with other environmental mycobacterial species and the vaccine strain *M. bovis* Bacille Calmette-Guerin (BCG). Such diagnostic tests will enable the transfer from skin testing regimes to vaccine regimes to be implemented.

A further advantage of the present invention is the provision of a test which can distinguish between those mammals that have been vaccinated against tuberculosis, and in particular *M. bovis,* and those which have been infected with *M. bovis*. This allows the selective slaughter of animals which would appear from current tests to all be infected, thereby saving the lives of many animals.

The applicants have also found that certain diagnostic reagents as described above, as well as polypeptides from which they are derived as well as some additional polypeptides, such as those shown in Figure 7 and SEQ ID NO 14 and 15 respectively, as well as variants and fragments thereof, produce a protective immune response, and therefore may be used as vaccines.

Thus the invention further provides a polypeptide comprising any one of SEQ ID NO 1, 2, 3, 4, 5 or 6,or variants thereof, or fragments of any of these, which produce a protective immune response in a mammal to whom they are administered, for use as a vaccine.

In addition, the present invention provides a polypeptide derived from an RD2 or RD14 region of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum* genomes, or a variant thereof, or a fragment of any of these, for use as a medicament, with the proviso that the polypeptide is not a MPT-64 polypeptide or a polypeptide encoded by the Rv1984c region of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum* genomes.

These polypeptides, or variants or fragments, are preferably used as a vaccine against tuberculosis caused by *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum.*

Suitably the polypeptide of the invention is derived from the *Mycobacterium tuberculosis* genome.
In particular the polypeptide of the invention comprises the sequence shown in SEQ ID Nos 14 or 15, or a variant thereof or fragment thereof. Most preferably the polypeptide is of SEQ ID NO 14 or 15 or an epitopic fragment thereof.

In particular, the invention provides a polypeptide comprising a fusion of a region of SEQ ID NO 14 and a region of SEQ ID NO 15, which fusion polypeptide is able to produce a protective immune response in a mammal to which it is administered. Particular examples of fusion polypeptides are illustrated hereinafter, as SEQ ID NOs 18, 20 and 22, and these, together with protective variants and fragments thereof form preferred embodiments of the invention.

Polypeptides which are protective are protective against tuberculosis infection and therefore may be used as a prophylactic or therapeutic vaccine, and these form a further aspect of the invention.

Therefore, particular vaccines comprise a polypeptide derived from an RD2 or RD14 region of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum* genomes, or a variant thereof or a fragment of any of these, which polypeptide produces a protective immune response against tuberculosis infection in a mammal to which it is administered, with the proviso that the polypeptide is not a MPT-64 polypeptide or a polypeptide encoded by the Rv1984c region of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum* genomes.

The vaccine is preferably used to vaccinate against tuberculosis. It may be used as a vaccine against tuberculosis in humans, cattle and other mammals including guinea pigs, badgers, possums and deer. It is, however, preferably used as a vaccine in cattle.

Preferably, the vaccine comprises one or more protein subunits.

Alternatively, it may comprise a nucleic acid such as a DNA or cDNA encoding for the protein or protein subunits.

When it comprises a nucleic acid, this is suitably incorporated into an expression vector, in such as way that the protein subunit is expressed in the host animal. For example, the nucleic acid may be incorporated into a virus vector such as a vaccinia or adenovirus vector, or a plasmid to form a so-called "naked DNA" vaccine. The vector may contain the usual expression control functions such as promoters, enhancers and signal sequences, as well as selection markers in order to allow detection of successful transformants. The nature of these will depend upon the precise nature of the vector chosen and will be known to or readily determinable by a person skilled in the art.

Preferably, vaccine compositions will further comprise an adjuvant such as in order to enhance the immune response to the biologically active material administered. Suitable adjuvants include pharmaceutically acceptable adjuvants such as Freund's incomplete adjuvant, aluminium compounds and, preferably adjuvants which are known to up-regulate mucosal responses such as CTB, the non-toxic pentameric B subunit of cholera toxin (CT).

According to a further aspect of the present invention, there is provided a nucleic acid encoding a polypeptide of the invention or a fragment or variant thereof. The nucleic acid may be DNA or RNA and where it is a DNA molecule, it may comprise a cDNA or

genomic DNA. These nucleic acids may themselves be useful as vaccines.

Preferably, the nucleic acids of the present invention are those shown as SEQ ID Nos 59 and 60, as well as or a variant or fragments thereof. One such variant, which encodes a fusion is SEQ ID NO 17.

According to a yet a further aspect of the invention, there is provided a pharmaceutical or veterinary composition comprising a protective polypeptide as described above, or a nucleic acid which encodes this, in combination with a pharmaceutically or veterinarily acceptable carrier.

The carriers may be solid or liquid as understood in the art. They may be obtained by conventional procedures using conventional pharmaceutical excipients, well known in the art.

In particular, the compositions of the invention may be in a form suitable for oral use (for example as tablets, lozenges, hard or soft capsules, aqueous or oily suspensions, emulsions, dispersible powders or granules, syrups or elixirs), for administration by inhalation (for example as a finely divided powder or a liquid aerosol), for administration by insufflation (for example as a finely divided powder) or for parenteral administration (for example as a sterile aqueous or oily solution for intravenous, subcutaneous, intramuscular or intramuscular dosing or as a suppository for rectal dosing.

The pharmaceutical or veterinary compositions are preferably in the form of a sterile injectable aqueous or oily suspension, which may be formulated according to known procedures using one or more of the appropriate dispersing or wetting agents and suspending agents, which have been mentioned above. A sterile injectable preparation may also be a sterile injectable solution

or suspension in a non-toxic parenterally-acceptable diluent or solvent, for example a solution in 1,3-butanediol.

Where the compositions of the invention comprise a nucleic acid, they are preferably formulated for parenteral administration and in particular intramuscular injection, although other means of application are possible as described in the pharmaceutical literature, for example administration using a Gene Gun, (Bennett et al., (2000), Vaccine 18, 1893-1901). Oral or intra-nasally delivered formulations are also possible. Such formulations include delivery of the plasmid DNA via a bacterial vector such as species of Salmonella or Listeria (Sizemore et al (1997). Vaccine 15, 804-807).

Formulation techniques generally are well known and are described for example in Chapter 25.2 in Volume 5 of Comprehensive Medicinal Chemistry (Corwin Hansch; Chairman of Editorial Board), Pergamon Press 1990.

The amount of active ingredient that is combined with one or more excipients to produce a single dosage form will necessarily vary depending upon the host treated and the particular route of administration. Dosage unit forms will generally contain about 1 mg to about 500 mg of an active ingredient.

The size of the dose for therapeutic or prophylactic purposes of the composition of the invention will naturally vary according to the age and sex of the animal or patient and the nature of the active component and the route of administration, according to well known principles of medicine. Generally speaking however, for administration to a human as a prophylactic vaccine, dosage units of from 0.25 µg to 2.5mg will be typically employed.

In yet a further aspect, the invention provides a method of protecting a mammal against infection against *Mycobacterium*

bovis, Mycobacterium tuberculosis or *Mycobacterium africanum* comprising administering to said mammal a polypeptide, a nucleic acid or a composition as described above. Where the polypeptide is a pharmaceutical or veterinary composition, the polypeptide may be administered directly. Alternatively, a nucleic acid encoding the polypeptides is administered to a mammal in a form in which it is expressed *in situ*.

According to another aspect of the present invention, there is provided a peptide derived from an RD2 or RD14 region of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum* genomes, or a variant thereof, or a fragment of any of these, which peptide produces a protective immune response against tuberculosis infection in a mammal to which it is administered.

Also encompassed by the present invention are peptides derived from the polypeptides of the present invention. Such peptides are preferably synthetic peptides.

According to yet another aspect of the present invention, there is provided the use of a polypeptide or peptide according to the present invention in the preparation of a vaccine.

The polypeptide may for example, have the sequence shown in SEQ ID NO. 7 or an epitopic fragment thereof.

The vaccine is preferably used to vaccinate against tuberculosis. It may be used as a vaccine against tuberculosis in both humans and cattle. It is, however, preferably used as a vaccine in cattle.

Preferably, the vaccine comprises protein subunits. Alternatively, it may comprise DNA or cDNA encoding for the subunits.

According to a further aspect of the present invention, there is provided a method of protecting a mammal against infection by *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum* comprising administering to said mammal a polypeptide, peptide or pharmaceutical or veterinary composition according to the present invention which produces an immune response against *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum.*

Cattle models of *M. bovis* infection and BCG vaccination were studied to identify highly immunogenic antigens from three genomic regions deleted in BCG Pasteur (RD1, RD2, RD14) [Behr, 1999 supra., Mahairas, 1996 supra] that could be useful as specific diagnostic reagents or subunit vaccine candidates. Five hundred and thirty six overlapping synthetic peptides derived from the sequence of 13 antigens (open reading frames) encoded in these regions were synthesised and used to diagnose infected or vaccinated cattle. The previously mentioned ESAT-6/CFP-10 peptide cocktail was also included as a gold standard with which to compare and all tests performed used the BOVIGAM ELISA for the detection of bovine IFN-γ.

The present invention will now be described only by way of example, in which reference shall be made to the Figures, in which:

Figure 1 shows the recognition of RD1 products by a *M. bovis* infected cow (A, C and E) and a PPD-A reactor (B, D and F). Whole blood was cultured in the presence of peptide pools of between 8-11 peptides representing RD1 (A and B), RD2 (C and D) and RD14 (E and F) at 5μg each peptide/ml. Dashed horizontal lines indicate positive cut-off (OD_{450} values with antigens minus OD_{450} without antigens $\geq$ 0.1);

Figure 2 shows IFN-γ responses induced by RD region antigens by *M. bovis* infected (22), BCG vaccinated (6) and PP-A reactor

cattle (10). Only the results of the pool/antigen stimulating the greatest IFN-γ response are shown. Green squares represent *M. bovis* infected cattle, red triangles represent PP-A reactors and blue circles represent BCG vaccinated cattle. Dashed horizontal line indicate the positive cut0off (OD_{450} values with antigens minus OD_{450} without antigens ≥ 0.1).

Figure 3 shows IFN-γ secretion induced by individual peptides from pool 3 (A) and pool 26 (B) in whole blood cultures from two representative animals. Whole blood was collected from *M. bovis* experimentally infected cattle and incubated for 48hrs with peptides (25ug/ml each). Results are expressed as delta mean optical density OD_{450} values with antigens minus OD_{450} without antigens) of duplicate determinations, with a positive cut-off of 0.1.

Figure 4 shows the antigens selected for evaluation.

Figure 5 shows the most frequently recognised antigen.

Figure 6 shows the sequence homology between peptide 3.2 from Rv3873 (shown as SEQ ID NO. 7) with other mycobacterial proteins.

Figure 7 shows the amino acid sequences which correspond to the open reading frames Rv1979c, Rv1769c, Rv1986, Rv3872, Rv3878, Rv1983, Rv3873 and Rv3879c which are shown as SEQ ID Nos 1 to 6, the amino acid sequences of the antigens which are particular vaccine candidates and in particular the Rv 1979 and Rv1769 antigens which are shown as SEQ ID Nos 14 and 15 respectively.

Figure 8 shows the nucleotide sequences of the Rv1979c, Rv1769c, Rv1986, Rv3872, Rv3878, Rv1983, Rv3873 and Rv3879c antigens whose coding sequences are shown as SEQ ID Nos 8 to 13, and whose genomic sequences are illustrated as SEQ ID Nos 61-66 respectively.

Figure 9 shows a diagnostic cocktail derived from SEQ ID NO 5.

Figure 10 shows a diagnostic cocktail derived from SEQ ID NO 3.

Figure 11 shows a diagnostic cocktail derived from SEQ ID NO 6.

Figure 12 shows the coding nucleotide sequences of the Rv 1979 and 1769 antigens which are shown as SEQ ID Nos 59 and 60, respectively, together with the genomic sequences 70 and 71 respectively.

Figure 13 shows the nucleotide coding and amino acid sequences of a novel vaccine as SEQ ID NOs 17 and 18 respectively.

Figure 14 shows the first half of a fusion insert from the ORF of Rv 1979c (SEQ ID NO 15) and the position with the ORF of the segment to be fused in the vaccine (bold).

Figure 15 shows the second half of a fusion insert from the ORF of Rv 1769 (SEQ ID NO 16) and the position with the ORF of the segment to be fused in the vaccine (bold).

Example 1

Diagnostic tests

The following results demonstrate that six antigens showed promise as diagnostic antigens with regard to their specificity, and that two more could be considered as potential vaccine candidates because they were highly immunogenic in all groups assayed.

MATERIALS AND METHODS

Cattle. Ca. 6 months old calves (Friesian or Friesian crosses) were obtained from herds free of bovine tuberculosis.

The following groups of cattle were used in this study:

M. bovis infection. Nine calves were infected with a GB *M. bovis* field strain from (AF 2122/97) by intratracheal instillation of 2 x 10^4 CFU as described [Buddle, B., et al, 1995. Vaccine 13: 1123-30; Buddle, B., et al. 1995. Res. Vet. Sci. 59: 10-6; Rhodes SG, et al. 2000. Infect. Immun 68:2573-2578]. Twelve calves were infected with an *M. bovis* field strain, isolated from a New Zealand infected cow using also intratracheal instillation (5 x 10^3 CFU). Bovine tuberculosis was confirmed in these animals by the presence of visible lessions in lymph nodes and lungs found at post-mortem examinations, by the histo-pathological examination of lesioned tissues and the culture of M. bovis from tissue samples collected from lymph nodes and lungs. Heparinised blood samples were obtained between 14-20 weeks after infection when strong and sustained *in vitro* tuberculin responses were observed. Data from a total of 21 experimentally infected cattle are presented in this study. One naturally infected animal was also used included in this group.

BCG vaccination. Calves were vaccinated with BCG Pasteur by subcutaneous injection of 10^6 CFU into the side of the neck followed 8 weeks later by a booster injection using the same route and dose [Buddle, 1995 supra.; Vordermeier, H. M., et al. 1999. Clin. Diagn. Lab. Immunol. 6:675-682]. Heparinised blood samples were taken between 4-6 weeks after the booster vaccination. Data from 6 calves will be presented in this study.

Uninfected controls. Heparinised blood from tuberculin skin test-negative calves from herds free of BTB (10 animals) was also obtained. These animals produced IFN-γ *in vitro* after stimulation with tuberculin from *M. avium* indicating that they have been exposed to environmental mycobacteria.

Antigens and peptides

Antigens: Bovine (PPD-B) and avian (PPD-A) tuberculins were obtained from the Tuberculin Production Units at the Veterinary Laboratories Agency-Weybridge and used in culture at 10 μg/ml.

Peptides: A set of five hundred and fifty two synthetic peptides spanning 13 open reading frames (20 residues long with a 12 residue overlap) was prepared by Multi-rod peptide synthesis. These were used in mapping experiments in pools of 10 peptides at 5μg each peptide/ml and 25μg/ml when used individually. The peptides were purchased from Chiron Mimotopes (Clayton, Australia). ESAT-6 and CFP-10 derived peptides were synthesised by solid phase peptide synthesis and formulated into a peptide cocktail as described earlier [Vordermeier, H. M. et al. 2001. Clin. Diagn. Lab. Immunol. 8:571-8]. They were also used at 5μg each peptide/ml. Peptide purity and sequence fidelity of ESAT-6 and CFP-10 derived peptides was confirmed by analytical reverse-phase HPLC and by electron-spray mass spectrometry, respectively.

Interferon-gammaELISA. Whole blood cultures were performed in 96-well plates in 0.2ml/well aliquots by mixing 0.1 ml of heparinised blood with an equal volume of antigen containing solution [Vordermeier, 1999 supra.]. Supernatants were harvested after 48 h of culture at 37°C/5% CO_2 in a humidified incubator. Interferon-gamma (IFN-γ) concentration was determined using BOVIGAM™ ELISA kit (Biocore AH, Omaha, NE). Results were deemed positive when the OD_{450} [PPD-B] minus OD_{450} values with antigens minus OD_{450} value without antigens were $\geq$ 0.1. For comparative analysis of PPD-B vs. PPD-A responses, a positive result was defined by an OD_{450} [PPD-A $\geq$0.1, and OD_{450} [PPD-B] minus OD_{450} [unstimulated] $\geq$0.1.

Bioinformatics

The DNA sequence of *M. tuberculosis* H37Rv was visualised using either the ARTEMIS display tool [Rutherford, K., J. et al. 2000.

Artemis: Sequence visualisation and annotation. Bioinformatics B10: 944 - 5] or the TubercuList database (http://genolist.pasteur.fr/TubercuList/) BLAST searches were performed from within TubercuList, or using the NCBI BLAST server (http://www.ncbi/nlm.nih.gov/BLAST)

RESULTS

Selection of candidate antigens from the RD1, RD2, and RD14 regions of *M. bovis*

Thirteen ORFs from the RD1, RD2 and RD14 regions of *M. bovis* were selected for screening. These regions are deleted in BCG Pasteur and proteins encoded within these regions hold promise as candidate antigens for the differential diagnosis of *M. bovis* infected animals from BCG vaccinated cattle and as potential vaccine candidates. Selection criteria were that the ORF should encode a protein that either (i) showed no, or minimal, sequence similarity to other proteins in *M. tuberculosis* or other organisms, (ii) belonged to the PE or PPE protein family, (iii) had the potential of being induced or upregulated *in vivo* (e.g. amino acid transporters), or (iv) had the potential to be secreted. The designations of the antigens encoded by the selected ORF (Rv number), their sizes, and putative functions are listed in Figure 4.

Immunogenicity of selected antigens in *M. bovis* infected, BCG vaccinated and environmentally sensitised cattle

Five hundred and thirty six overlapping peptides derived from the sequences of these antigens were synthesised. Peptides were then formulated into pools of approximately 10 neighbouring overlapping peptides, which resulted in 52 peptide pools. Figure 4 indicates the pool in relation to the antigens they represent as well as the total number of peptides/antigen required to ensure complete sequence coverage. Blood samples were obtained from 22 *M. bovis* infected animals, 6 *m. bovis* BCG Pasteur vaccinated animals and 10 un-vaccinated/un-infected

controls: Whole blood cultures in the presence of PPD-B, PPD-A, peptide pools and a cocktail of 10 synthetic peptides derived from ESAT-6 and CFP-10, were established and the amount of IFN-γ determined after 48 h of culture.

As expected, all *M. bovis* infected and BCG vaccinated animals responded more strongly to bovine tuberculin PPD-B than to avian tuberculin PPD-A (median responses and range: *M. bovis* infected: PPD-B=1.593(0.274-3.500), PPD-A=1.313(0.066-3.455)); BCG vaccinated: PPD-B=0.886(0.181-2.244), PPD-A=0.5115(0.274-2.234)); Uninfected, non-vaccinated control against animals responded strongly to avian PPD (PPD-A) indicating that they were sensitised by environmental mycobacteria (Median responses and ranges: PPDB=0.230(0.090-0.684), PPD-A=0.686(0.162-1.822)); they will be described hereinafter as *PPD-A reactors*. Next the immunogenicity of the peptide pools described in Figure 4 was assessed. Figure 1 depicts the results obtained with blood from two representative animals, one infected with *M. bovis*, the other a PPD-A reactor. The *M. bovis* infected animal recognised at least one peptide pool from each antigen (Figure 1A, C, E), indicating that cellular responses were induced after *M. bovis* infection against all 13 antigens selected. In contrast, none of the peptide pools induced IFN-γ secretion in whole blood from the environmentally sensitised PPD-A reactor (Figure 1B, D F).

The peptide-induced IFN-γ responses of all 38 M. bovis infected, BCG vaccinated and PPD-A-reactors (uninfected controls) to the 13 antigens are summarised in Figure 2. When antigens were covered by more than one peptide pool, the result of the pool stimulating the most IFN-γ secretion is shown. Interestingly, all 13 antigens were recognised by *M. bovis* infected cattle all be it with the percentage of responding cattle (responder frequencies) varying between 21 and 86 %. The most frequently recognised antigens were Rv3873, Rv3879c and Rv1769, with responder frequencies of 82, 77 and 86% respectively, whereas Rv1984c and Rv1772 were recognised only by 21 and 36% of

infected calves. Interestingly, several of the most prominently recognised antigens were members of the PE/PPE protein family (e.g., Rv3873, with a responder frequency of 82%).

Surprisingly, considering the absence of the genes encoding these antigens in BCG Pasteur, 9 of the 13 antigens tested, stimulated a positive response in BCG vaccinated animals (Rv3873, Rv3879c, Rv1979c, Rv1983, Rv1987, Rv1989c, Rv1768, Rv1769 and RV 1772, with a range in responder frequencies of 17 - 100%). The remaining four antigens were recognised by *M. bovis* infected cattle only (Rv3872, Rv3878, Rv1984c, and Rv1986, with a range of responder frequencies of 21-59%). The responder frequencies of the 8 most immunogenic antigens are summarised in Figure 5. In addition, 21/22 *M. bovis* infected animals responded to a previously characterised peptide cocktail derived from CFP-10 and ESAT-6 (reference) that had been included for comparison (median responses and range: 1.281 (0.011-2.825)).

EXAMPLE 2

The combination of antigens offers improved sensitivity

It is unlikely that a single diagnostic antigen, however specific, could impart enough sensitivity to provide population coverage; therefore combinations of specific antigens will be needed. It was therefore determined whether such antigen combinations could improve test sensitivity. Two scenarios were considered: firstly, antigens suitable for differential diagnosis, i.e., not recognised by BCG vaccinated animals or PPD-A reactors. The three antigens most frequently recognised by *M. bovis* infected animals fulfilling this criteria are Rv1986, Rv3872, and Rv3878 (Figure 5). Combining their results indicated that 82% of the infected animals would have been correctly identified by their responses to either of these three antigens (Figure 5).

Secondly, we considered the three most immunodominant antigens (Rv1983, Rv3873, Rv3879c) that were not recognised by PPD-A

reactors, but were recognised by BCG vaccinated calves, i.e., antigens capable of distinguishing between *M. bovis* infection and animals sensitised by environmental mycobacteria for example, *M. avium* (specific diagnosis). Taken together, these antigens would have identified 20/22 (91%) of the *M. bovis* infected animals (Figure 5). Interestingly, if Rv3878 from the first category was considered together with Rv3873 and Rv3879c from this category. 21/22 (95%) of the *M. bovis* infected animals would have been detected (Figure 5).

EXAMPLE 3

Responses of peptide pools can be the result of a single peptide

The peptide pools formulated contain between 8-11 peptides (see Figure 4 for details of peptide pools). To determine whether IFN-γ responses of pools were due to single or multiple peptide constituents, the individual peptides of pool 3 (representing residues 89-188 from Rv3873) and pool 26 (representing residues 161-252 from Rv1983) were tested using blood from 5 *M. bovis* infected animals. All three animals tested that recognised pool 3 responded exclusively to peptide 3.2 (residues 97-116 - SEQ ID NO 7), whereas both animals tested that responded to pool 26 only recognised peptide 26.2 (residues 169-188). The results shown in Figure 3 give results from one representative animal responding to pools 3 (Fig. 3A) or 26 (Fig.3B), respectively. These data suggest that the individual peptides imparting antigenicity can be identified from immunodominant pools and that pool immunogenicity can be attributed to single peptides.

The effective use of comparative genomics in combination with synthetic peptides to identify and screen thirteen potential antigens encoded by ORFs located in the RD1, RD2, and RD14 regions of the *M. tuberculosis* has been demonstrated. These results indicated that six antigens in particular showed promise as diagnostic antigens because they were either (i) recognised by *M. bovis* infected animals alone, but not by BCG vaccinated or controls (differential diagnosis, Figure 5) or (ii) by infected

animals and vaccinated animals but not by environmental mycobacteria exposed controls (specific diagnosis, Figure 5).

In general, all 13 antigens tested were recognised with responder frequencies varying between 21 and 86%. It is likely that a combination of several factors determines whether and to what degree mycobacterial proteins are immunogenic after infection. These factors could include (a) parameters intrinsic to the bacterium, such as the abundance of the protein, its sub-cellular location, post-translational modification, participation in macromolecular complexes, and *in vivo* regulation; and (b) factors relating to the immune system, including location of the antigen with respect to the phagosome, proteolytic sensitivity, and the presence of motifs suitable for interaction with TAP transporters and different MHC alleles within the antigen.

The present invention exploits the use of pools of overlapping synthetic peptides derived from the sequences of these proteins. In a pilot experiment where the peripheral monocyte blood cell (PBMC) was isolated from 8 cattle experimentally infected with *M. bovis* and stimulated them with either recombinant ESAT-6 or a cocktail of 11 synthetic peptides spanning the whole sequence of ESAT-6, it was concluded that the numbers of IFN-γ producing cells, determined in this case by ELISPOT, demonstrated equivalent responses to recombinant protein and synthetic peptides ($r=0.92$, $p<0.0001$). The number of peptide pools that represent the sequences of each ORF varies depending on the size of the antigen, as illustrated in Figure 4. It was demonstrated that the combined results from Rv3873, Rv3878 and Rv3879c resulted in an overall responder frequency of 95%. These 3 antigens are represented by a total of 16 different peptide pools, containing 169 individual peptides. However, the same frequency of recognition can be obtained using just 3 pools out of the 16 pools assayed (pools 3, 8 and 9), i.e., 30 peptides, suggesting the presence of the immunodominant epitomes within

these three pools. Indeed, the number of peptides needed to achieve responder frequencies similar to that with the complete set of overlapping peptides could even be significantly lower since the data described in Figure 3 demonstrates that only one or two immunodominant peptides can be responsible for the immunogenicity of the whole pool. If these peptides were to be recognised promiscuously in the context of multiple MHC molecules, as has been described in the recognition of other mycobacterial antigens by human, murine and bovine CD4+ T cells [Vordermeier, H. M., et al. 2000. Clin. Infect. Dis. 30:S291-S298; Vordermeier, 2001 supra.; Vordermeier, H. M. 1995. Eur. Respir. J. Suppl. 20:657s-667s; Lightbody, K. A., et al 1998. Scand. J. Immunol. 48: 44-51; Lightbody, K. A., et al. 1998. Immunology 93:314-22; Pollock, J. M., et al.1994. Immunology. 82:9-15; Pollock, J. M., et al. 1995. Scand. J. Immunol. 41:85-93.], the number of peptides required to achieve wide population coverage could be relatively low as has been demonstrated before for ESAT-6 and CFP-10 derived peptides [Vordermeier, 2001 supra., Lalvani, A., et al. 2001. J. Infect. Dis. 183:469-477; Lalvani, A., et al. 2001. . Am. J. Respir. Crit. Care. Med. 163: 824-8].

Peptides as immuno-diagnostic reagents can therefore constitute a practical alternative to recombinant proteins, in addition to substituting them as reagents to assess immunogenicity. The fact that all three animals tested, two from the UK and one from New Zealand, recognised the same peptide within pool 3 (peptide 3.2) is encouraging in this context.

Interestingly, the previously described peptide cocktail containing peptides derived from ESAT-6 and CFP-10 was also recognised by 95% of the *M. bovis* infected animals tested, in fact the same animals that responded to the combination of Rv3873, Rv3878 and Rv3879c.

As described in Figure 6, 4 PPD/ PE genes were selected for testing (Rv3872, Rv3873, Rv1983 and Rv1768) and gave responder frequencies of between 45-82% when assayed in the *M. bovis* infected cattle. Little is known about the function or immunogenicity of these proteins, which account for approximately 10% of the total coding capacity of the *M. tuberculosis* genome.

As described in Figure 3, peptide 3.2 is a highly immunogenic component of pool 3 derived from the sequence of Rv3873, a member of the PPE family of proteins. The pool consistently produces positive responses when assayed in *M. bovis* infected cattle with a responder frequency of 82% but was also recognised in BCG vaccinated animals. This is a surprising outcome given that its gene is deleted in BCG and that no homologous proteins were found elsewhere in the BCG genome. However, the unit of cross-reactivity is the epitope, less than 20 amino acids long, that is recognised by T cells in the context of MHC molecules. Consequently, the molecular nature of cross-reactivity can only be addressed once these epitopes have been identified.

Therefore we used the sequence of peptide 3.2 (shown as SEQ ID NO.9) to search for similar regions with other genes found within the *M. tuberculosis* genome.

Figure 6 shows the results using the Basic Local Alignment Search Tool (BLAST) program [NCBI. Basic Local Alignment Search Tool (BLAST). http://www/ncbi.nlm.nih.gov/BLAST/] to identify similarity between mycobacterial proteins. The table shown in Figure 6 highlights several sequences that contain amino acid identities of greater than 50%. These include five proteins from the *M. tuberculosis* genome, all of which are also members of the PPE family and several others identified in proteins of various mycobacterial species. The peptide covers an area of the gene that encodes two motifs identified in a number of PPE family members during their annotation [Tekaia, F., et al 1999.

Tuber. Lung Dis. 79:329-42. TubercuList. MAST - Motif Alignment and Search Tool http://genolist.pastuer.fr/TubercuList/].

This suggests that the cross reactive nature of the peptide is a result of similarity with other PPE family members located elsewhere in the genome of *M. tuberculosis* and therefore the genome of *M. bovis* BCG Pasteur. We conducted BLAST searches for the other identified cross-reactive antigens (e.g., Rv1979c) by comparing the whole genes in steps of 20 amino acids, representing the corresponding peptides, and were able to find numerous similar amino acid sequences in other mycobacterial proteins outside the deleted regions.

The use of peptides instead of recombinant proteins, has several advantages already discussed. However, with regard to the observed cross reactivity of antigens between BCG vaccinated and *M. bovis* infected animals, this peptide-based approach has other distinct advantages. If ORF Rv1987 is taken as an example, it appears unsuitable as a differential diagnostic reagent due to the high cross reactivity in the BCG vaccinated cattle. However, the responder frequency of 57% in *M. bovis* infected cattle is due to the recognition of two pools with responder frequencies of 47% and 53% respectively. Whilst one pool is recognised by 50% of the BCG vaccinated animals, the other is not recognised. Therefore, the diagnostic potentials of this antigen can still be realised by using only peptides derived from the second peptide pool.

In summary, therefore, the analysis of peptides, derived from genes deleted in BCG Pasteur, has led to the identification of antigens for diagnosis and even vaccination. In particular, antigens that can form the basis of diagnostic reagents to either differentiate between infected and BCG vaccinated animals or to improve the specificity of PPD *per se* are described. In addition, it has also been demonstrated for the first time that members of the both the PE and PPE families of proteins induced

cellular immune response after mycobacterial infection of a target species.

EXAMPLE 4

Immunogenic antigens

The Experiment described in Example 1 above could also be used to demonstrate that two antigens could be considered as potential vaccine candidates because they were highly immunogenic in all groups assayed.

RESULTS

The immunogenicity of the peptide pools described in Figure 4 was assessed. 2 antigens, Rv1979c and Rv1769 (of SEQ ID NO 14 and SEQ ID NO 15 respectively) showed responder frequencies of 73% and 86% respectively in the 22 cattle experimentally infected with *M. bovis* (Figure 5). These two antigens were also strongly recognised by BCG vaccinated cattle with responder frequencies of 67 and 100% respectively. Uniquely however, they also showed 40 and 30% responder frequency in the PPD-A.

These results indicated that two antigens can be considered vaccine candidates since they were recognised by T cells from all 3 categories (Rv1979c and Rv1769).

EXAMPLE 5

DNA Fusion Vaccine

A DNA vaccine comprising a fusion of two internal gene sequences was constructed in the vector pvmcLINK (Vordermeier et al. Vaccine 2000 Dec 8;19(9-10):1246-55).

The two gene sequences, derived from a section of the sequence of Rv1979c and Rv1769 respectively, were generated by polymerase chain reaction (PCR) and ligated together via their restriction enzyme (RE) digested termini. The two sections encode for polypeptides that stimulate the production of gamma interferon in blood from cattle inoculated with mycobacterium bovis,

mycobacterium bovis BCG and others exposed to environmental mycobacteria (Cockle et al. Infect Immun 2002 Dec;70(12):6996-7003). The vector itself was then cut using RE's at a specific position and the fused insert sequence ligated within it.

The complete DNA sequence of the DNA fusion vaccine is shown in Figure 13. The sequence highlighted in bold is that of the DNA fusion insert comprised of two ORF sections from Rv1769 and Rv1979c respectively. The rest of the sequence is that of the vaccine vector pvmcLINK.

The section of the open reading frame (ORF) from Rv1979c that has been fused into the DNA vaccine is shown in Figure 14. The section is an internal gene sequence that encodes an area of the gene that, when assayed in the translated form as polypeptides, stimulates the production of gamma interferon in blood from cattle inoculated with mycobacterium bovis, mycobacterium bovis BCG and others exposed to environmental mycobacteria (Cockle et al. supra.). The position of the antigenic gene section within the ORF is highlighted in the sequence in bold.

Figure 15 shows the section of the open reading frame (ORF) from Rv1769 that has been fused into the DNA vaccine. The section is a gene sequence that encodes an area of the gene that, when assayed in the translated form as polypeptides, stimulates the production of gamma interferon in blood from cattle inoculated with mycobacterium bovis, mycobacterium bovis BCG and others exposed to environmental mycobacteria (Cockle et al. supra). The position of the antigenic gene section within the ORF is highlighted in the sequence in bold.

All references mentioned in the above specification are herein incorporated by reference. Other modifications of the present invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although

the invention has been described in connection with the specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention, which are obvious to those skilled in the art, are intended to be within the scope of the following claims.

CLAIMS

1. A diagnostic reagent comprising a peptide comprising an epitope from at least one polypeptide selected from Rv1986 (SEQ ID NO 1), Rv3878 (SEQ ID NO 3), Rv 1983 (SEQ ID NO 4), Rv3873 (SEQ ID NO 5) or Rv3879 (SEQ ID NO 6).

2. A diagnostic reagent according to claim 1 which comprises a peptide comprising a series of consecutive amino acids from within the polypeptide sequences defined in claim 1.

3. A diagnostic reagent according to claim 1 or claim 2 which comprises an epitope from SEQ ID Nos 3, 5 or 6.

4. A diagnostic reagent according to claim 3 which comprises a peptide which include an epitope from SEQ ID NO 23 as shown in Figure 9, or a fragment thereof.

5. A diagnostic reagent according to claim 4 wherein the fragment is selected from SEQ ID NOs 7, 25, 28 and 29.

6. A diagnostic reagent according to claim 3 which comprises SEQ ID NO 23, SEQ ID NO 7, SEQ ID NO 25, SEQ ID NO 28 or SEQ ID NO 29.

7. A diagnostic reagent according to claim 3 which comprises an epitope from SEQ ID NO 35, shown in Figure 10.

8. A diagnostic reagent according to claim 7 which comprises SEQ ID NO 35 or a fragment thereof.

9. A diagnostic reagent according to claim 3 which comprises an epitope from SEQ ID NO 48, shown in Figure 11 hereinafter, or a fragment thereof.

10. A diagnostic reagent according to claim 9 wherein the fragment is of SEQ ID NO 51.

11. A diagnostic reagent according to claim 9 which comprises SEQ ID NO 48 or variant or fragment thereof,

12. A diagnostic reagent according to claim 11 which comprises SEQ ID NO 51.

13. A diagnostic kit comprising at least two diagnostic reagents, at least one of which is a diagnostic reagent according to any one of claims 1 to 12.

14. A diagnostic kit according to claim 13 which further comprises one or more polypeptides or peptides derived from ESAT-6 and/ or the CFP-10 polypeptides.

15. A diagnostic kit according to claim 13 wherein the diagnostic reagents are selected so that they are able to differentiate between *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum* -infected mammals and mammals vaccinated against *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum.*

16. A nucleic acid which encodes a diagnostic reagent according to any one of claims 1 to 12.

17. A method for diagnosing infection in a host, or exposure of a host, to a mycobacterium, said method comprising
i) contacting a population of cells from the host with a diagnostic reagent according to any one of claims 1 to 12; and
ii) determining whether the cells of said cell population recognise the diagnostic reagent.

18. A method according to claim 17 wherein the population of cells is a population of T-cells.

19. A polypeptide comprising any one of SEQ ID NO 1, 2, 3, 4, 5 or 6,or variants thereof, or fragments of any of these, which produce a protective immune response in a mammal to whom they are administered, for use as a medicament.

20. A polypeptide derived from an RD2 or RD14 region of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum* genomes, or a variant thereof, or a fragment of any of these, for use as a medicament, with the proviso that the polypeptide is not a MPT-64 polypeptide or a polypeptide encoded by the Rv1984c region of the *Mycobacterium tuberculosis, Mycobacterium bovis* or *Mycobacterium africanum* genomes.

21. A polypeptide according to claim 19 or claim 20 which is derived from the *Mycobacterium tuberculosis* genome.

22. A polypeptide according to claim 20 which comprises the sequence shown in SEQ ID Nos 14 or 15, or a variant thereof or fragment thereof.

23. A polypeptide according to claim 22 which is of SEQ ID NO 14 or 15 or an epitopic fragment thereof.

24. A polypeptide according to claim 20 which comprises a fusion of a region of SEQ ID NO 14 and a region of SEQ ID NO 15, which fusion polypeptide is able to produce a protective immune response in a mammal to which it is administered.

25. A polypeptide according to claim 24 which comprises SEQ ID NOs 18, 20 and 22, or a protective variant or and fragment thereof.

26. A vaccine comprising a polypeptide according to any one of claims 19 to 25.

27. A vaccine according to claim 26 comprising one or more protein subunits.

28. A nucleic acid which encodes a polypeptide according to any one of claims 19 to 27 for use as a vaccine.

29. A nucleic acid according to claim 28 which comprises SEQ ID Nos 59 or 60, or a variant or fragment thereof.

30. A nucleic acid according to claim 28 which comprises SEQ ID NO 17.

31. A pharmaceutical or veterinary composition comprising a protective polypeptide as described above, or a nucleic acid which encodes this, in combination with a pharmaceutically or veterinarily acceptable carrier.

32. A method of protecting a mammal against infection by *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum* comprising administering to said mammal a polypeptide according to any one of claims 19 to 25, a nucleic acid according to any one of claims 28 to 30 or a composition according to claim 31.

33. A method of protecting a mammal against infection by *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum* comprising administering to said mammal a polypeptide, peptide or pharmaceutical or veterinary composition according to the present invention which produces an immune response against *Mycobacterium bovis, Mycobacterium tuberculosis* or *Mycobacterium africanum.*

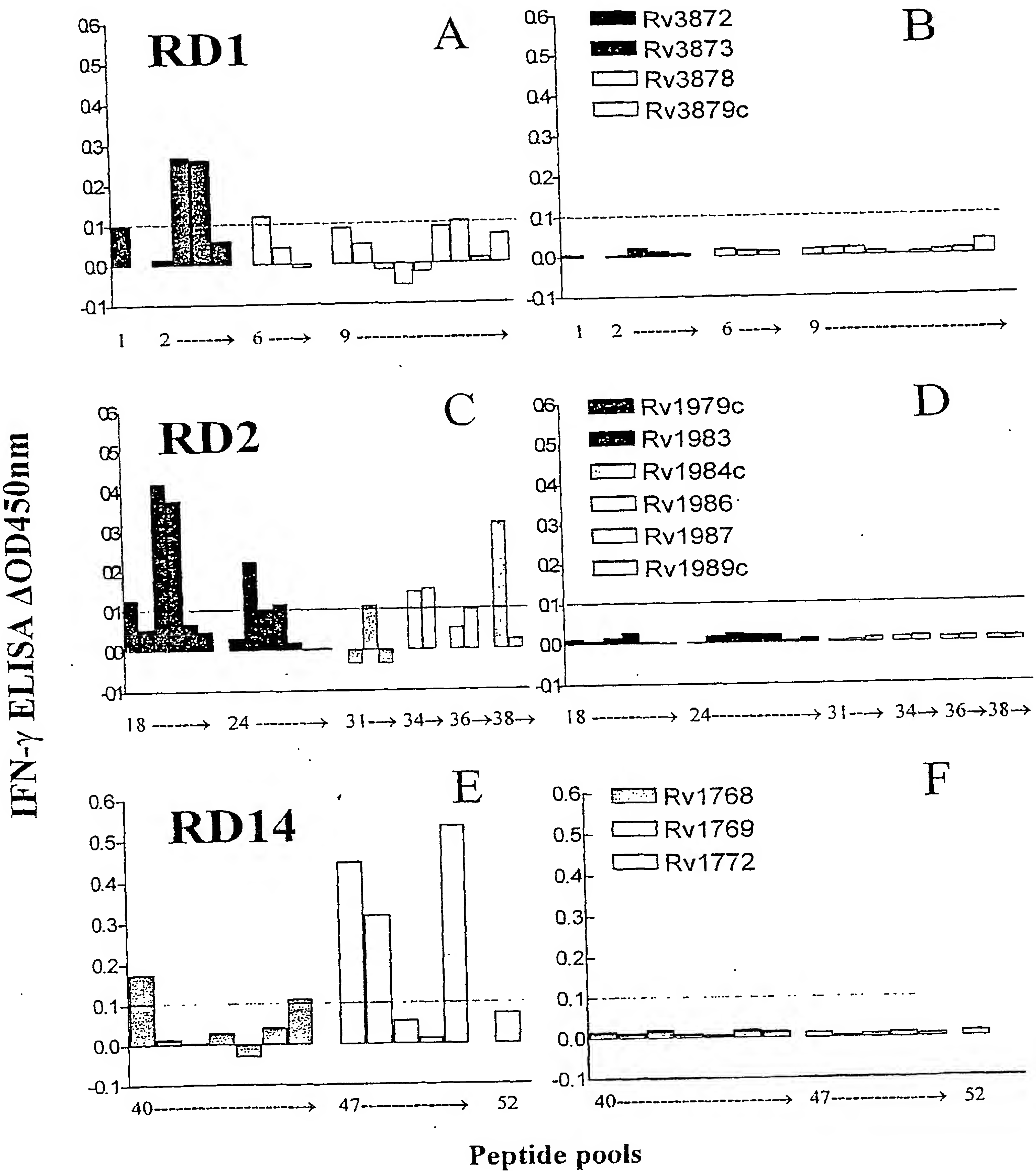
RD1
A
B
Rv3872
Rv3873
Rv3878
Rv3879c
RD2
C
D
Rv1979c
Rv1983
Rv1984c
Rv1986
Rv1987
Rv1989c
RD14
E
F
Rv1768
Rv1769
Rv1772
IFN-γ ELISA ΔOD450nm
Peptide pools

Figure 2

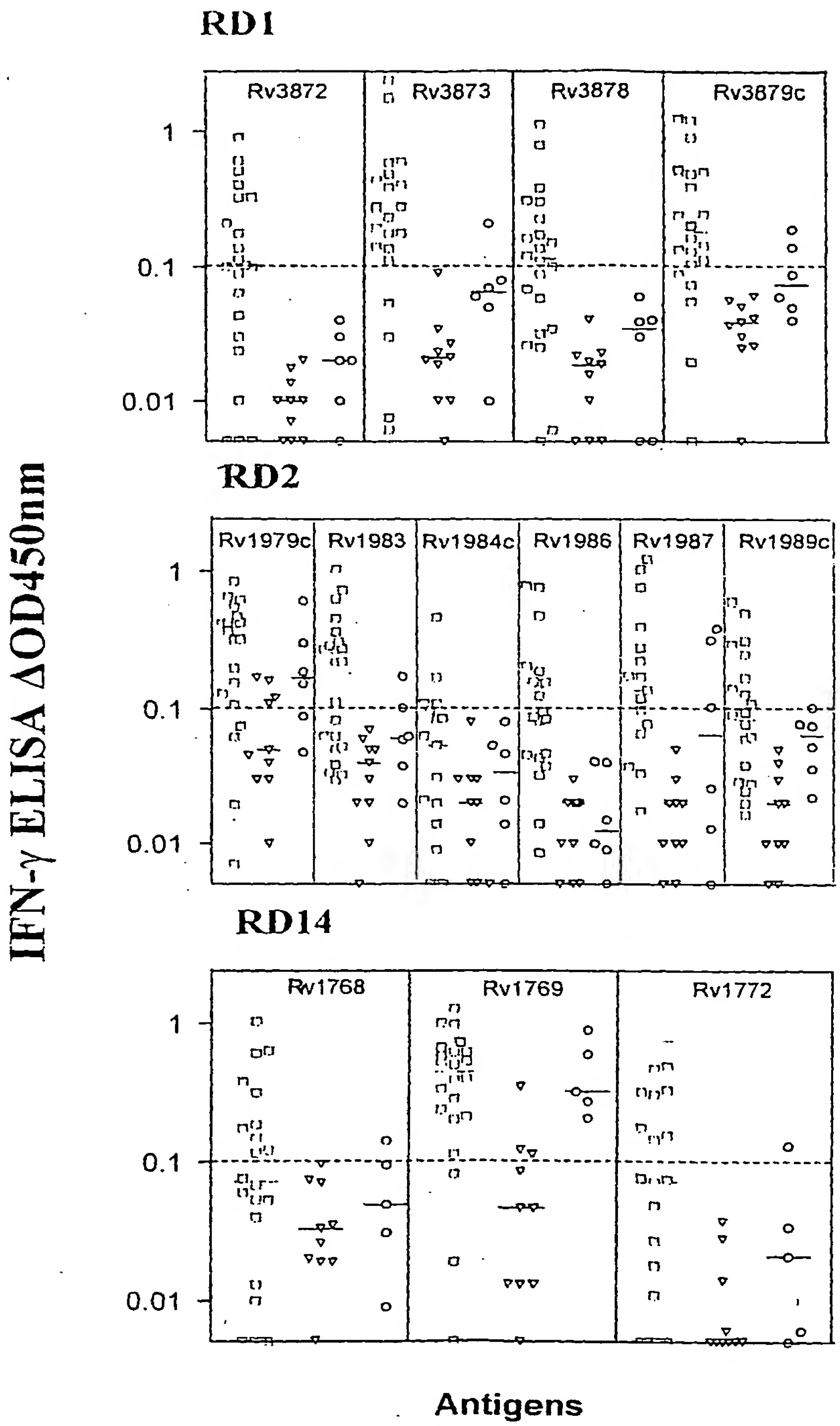
RD1
Rv3872
Rv3873
Rv3878
Rv3879c
1
0.1
0.01
RD2
Rv1979c
Rv1983
Rv1984c
Rv1986
Rv1987
Rv1989c
RD14
Rv1768
Rv1769
Rv1772
IFN-γ ELISA ΔOD450nm
Antigens

Figure 3

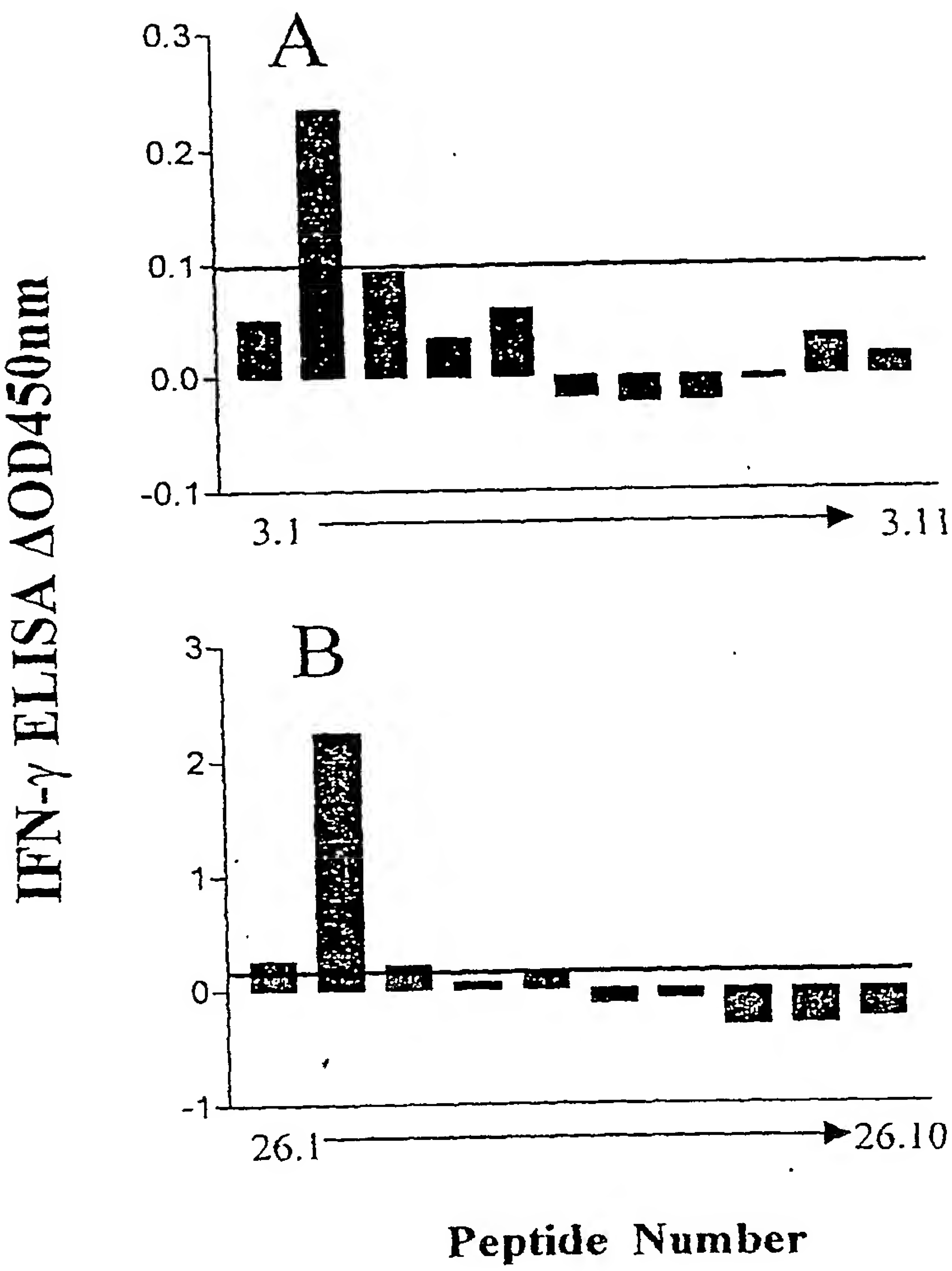
A
B
0.3
0.2
0.1
0.0
-0.1
3.1
3.11
3
2
1
0
-1
26.1
26.10
IFN-γ ELISA ΔOD450nm
Peptide Number

Figure 4. RD antigens selected for evaluation in this study

Deleted Region	Designation[a]	Size (Amino Acids)	Peptide Pools[b]	Putative Function[c]
RD1	Rv3872	99	1 (10)	Member of PE-like protein family
	Rv3873	368	2-5 (40)	Member of *M. tuberculosis* PPE family
	Rv3878	280	6-8 (30)	Unknown, alanine-rich protein
	Rv3879c	729	9-17 (90)	Unknown, alanine-proline-rich protein
RD2	Rv1979c	481	18-23 (60)	Possible amino acid permease
	Rv1983	558	24-30 (70)	Member of the PE-PGRS sub-family of glycine-rich proteins
	Rv1984c	217	31-33 (30)	Probable secreted cutinase
	Rv1986	199	34-35 (20)	Possible lysine transporter
	Rv1987	142	36-37 (20)	Possible chitinase
	Rv1989c	186	38-39 (20)	Unknown
RD14	Rv1768	618	40-46 (70)	Member of the PE-PGRS sub-family of glycine-rich proteins
	Rv1769	414	47-51 (50)	Similar to *Streptomyces coelicolor* hypothetical protein
	Rv1772	103	51-52 (20)	Unknown

[a]Rv designation of ORF as defined [Cole, 1998 1998. Nature **393**: 537-44]

[b]Number of peptide pools required to cover full sequence (total number of peptides required shown in brackets)

[c]Putative function as suggested [Cole, 1998 supra.]

Figure 5. List of most frequently recognised antigens[a]

Designation	Responder Frequency %			Potential Application
	[b]*M.bovis* Reactors	[c]BCG Vaccinated	[d]*M.avium* Reactors	
Rv1986	41	0	0	Differential Diagnostics
Rv3872	50	0	0	
Rv3878	59	0	0	
Combined	82	0	0	
Rv1983	59	33	0	Specific Diagnostics
Rv3873	82	17	0	
Rv3879c	77	33	0	
Combined	91	50	0	
Rv1979c	73	67	40	Vaccines
Rv1769	86	100	30	

[a]Only antigens recognised by >40% of *M.bovis* infected animals are listed

[b]Results from 22 cattle experimentally infected with *M.bovis*

[c]Results from 5-6 BCG vaccinated cattle

[d]Results from 10 environmental mycobacteria sensitised cattle

Figure 6. Sequence homology between peptide 3.2 from Rv3873 with other mycobacterial proteins

Designation[a]	Putative Function	Amino Acid Sequence[b]	SEQ ID NO
RV3873	***M.tuberculosis* PPE family**	**AMATTPSLPEIAANHIT**	**7**
Rv3021c	*M.tuberculosis* PPE family	ALAEMPTLPELAANHLT	67
Rv0286	*M.tuberculosis* PPE family	ALAAMPTLAELAANHVI	68
Rv3018c	*M.tuberculosis* PPE family	ALAEMPTLPELAANHLT	69
Rv0280	*M.tuberculosis* PPE family	AVAAMPTLVELAANHTL	70

The homology search was performed using the BLAST program. [a]Designation of *M. tuberculosis* proteins as described [Cole, 1998 supra.]. [b]The sequence in *M. tuberculosis* and *M. bovis* was found to be identical. Amino acid residues are shown in the one letter code. Non-identical residues are underlined.

Figure 7

>Rv1983: 558 aa - M. tuberculosis - **SEQ. I.D. NO. 4**
1 - VSFLVVVPEF LTSAAADVEN IGSTLRAANA AAAASTTALA AAGADEVSAA VAALFARFGQ
61 - EYQAVSAQAS AFHQQFVQTL NSASGSYAAA EATIASQLQT AQHDLLGAVN APTETLLGRP
121 - LIGDGAPGTA TSPNGGAGGL LYGNGGNGYS ATASGVGGGA GGSAGLIGNG GAGGAGGPNA
181 - PGGAGGNGGW LLGNGGIGGP GGASSIPGMS GGAGGTGGAA GLLGWGANGG AGGLGDGVGV
241 - DRGTGGAGGR GGLLYGGYGV SGPGGDGRTV PLEIIHVTEP TVHANVNGGP TSTILVDTGS
301 - AGLVVSPEDV GGILGVLHMG LPTGLSISGY SGGLYYIFAT YTTTVDFGNG IVTAPTAVNV
361 - VLLSIPTSPF AISTYFSALL ADPTTTPFEA YFGAVGVDGV LGVGPNAVGP GPSIPTMALP
421 - GDLNQGVLID APAGELVFGP NPLPAPNVEV VGSPITTLYV KIDGGTPIPV PSIIDSGGVT
481 - GTIPSYVIGS GTLPANTNIE VYTSPGGDRL YAFNTNDYRP TVISSGLMNT GFLPFRFQPV
541 - YIDYSPSGIG TTVFDHPA

>Rv1986: 199 aa - M. tuberculosis - **SEQ. I.D. NO. 1**
1 - VNSPLVVGFL ACFTLIAAIG AQNAFVLRQG IQREHVLPVV ALCTVSDIVL IAAGIAGFGA
61 - LIGAHPRALN VVKFGGAAFL IGYGLLAARR AWRPVALIPS GATPVRLAEV LVTCAAFTFL
121 - NPHVYLDTVV LLGALANEHS DQRWLFGLGA VTASAVWFAT LGFGAGRLRG LFTNPGSWRI
181 - LDGLIAVMMV ALGISLTVT

>Rv3872: 99 aa - M. tuberculosis - **SEQ. I.D. NO. 2**
1 - MEKMSHDPIA ADIGTQVSDN ALHGVTAGST ALTSVTGLVP AGADEVSAQA ATAFTSEGIQ
61 - LLASNASAQD QLHRAGEAVQ DVARTYSQID DGAAGVFAE

>Rv3873: 368 aa - M. tuberculosis - **SEQ. I.D. NO. 5.**
1 - MLWHAMPPEL NTARLMAGAG PAPMLAAAAG WQTLSAALDA QAVELTARLN SLGEAWTGGG
61 - SDKALAAATP MVVWLQTAST QAKTRAMQAT AQAAAYTQAM ATTPSLPEIA ANHITQAVLT
121 - ATNFFGINTI PIALTEMDYF IRMWNQAALA MEVYQAETAV NTLFEKLEPM ASILDPGASQ
181 - STTNPIFGMP SPGSSTPVGQ LPPAATQTLG QLGEMSGPMQ QLTQPLQQVT SLFSQVGGTG
241 - GGNPADEEAA QMGLLGTSPL SNHPLAGGSG PSAGAGLLRA ESLPGAGGSL TRTPLMSQLI
301 - EKPVAPSVMP AAAAGSSATG GAAPVGAGAM GQGAQSGGST RPGLVAPAPL AQEREEDDED
361 - DWDEEDDW

>Rv3878: 280 aa - M. tuberculosis - **SEQ. I.D. NO. 3**
1 - MAEPLAVDPT GLSAAAAKLA GLVFPQPPAP IAVSGTDSVV AAINETMPSI ESLVSDGLPG
61 - VKAALTRTAS NMNAAADVYA KTDQSLGTSL SQYAFGSSGE GLAGVASVGG QPSQATQLLS
121 - TPVSQVTTQL GETAAELAPR VVATVPQLVQ LAPHAVQMSQ NASPIAQTIS QTAQQAAQSA
181 - QGGSGPMPAQ LASAEKPATE QAEPVHEVTN DDQGDQGDVQ PAEVVAAARD EGAGASPGQQ
241 - PGGGVPAQAM DTGAGARPAA SPLAAPVDPS TPAPSTTTTL

>Rv3879c: 729 aa - M. tuberculosis - **SEQ. I.D. NO. 6**
1 - MSITRPTGSY ARQMLDPGGW VEADEDTFYD RAQEYSQVLQ RVTDVLDTCR QQKGHVFEGG
61 - LWSGGAANAA NGALGANINQ LMTLQDYLAT VITWHRHIAG LIEQAKSDIG NNVDGAQREI
121 - DILENDPSLD ADERHTAINS LVTATHGANV SLVAETAERV LESKNWKPPK NALEDLLQQK
181 - SPPPPDVPTL VVPSPGTPGT PGTPITPGTP ITPGTPITPI PGAPVTPITP TPGTPVTPVT
241 - PGKPVTPVTP VKPGTPGEPT PITPVTPPVA PATPATPATP VTPAPAPHPQ PAPAPAPSPG
301 - PQPVTPATPG PSGPATPGTP GGEPAPHVKP AALAEQPGVP GQHAGGGTQS GPAHADESAA
361 - SVTPAAASGV PGARAAAAAP SGTAVGAGAR SSVGTAAASG AGSHAATGRA PVATSDKAAA
421 - PSTRAASART APPARPPSTD HIDKPDRSES ADDGTPVSMI PVSAARAARD AATAAASARQ
481 - RGRGDALRLA RRIAAALNAS DNNAGDYGFF WITAVTTDGS IVVANSYGLA YIPDGMELPN
541 - KVYLASADHA IPVDEIARCA TYPVLAVQAW AAFHDMTLRA VIGTAEQLAS SDPGVAKIVL
601 - EPDDIPESGK MTGRSRLEVV DPSAAAQLAD TTDQRLLDLL PPAPVDVNPP GDERHMLWFE
661 - LMKPMTSTAT GREAAHLRAF RAYAAHSQEI ALHQAHTATD AAVQRVAVAD WLYWQYVTGL
721 - LDRALAAAC

>Rv1979c: 481 aa - M. tuberculosis - **SEQ I.D. NO. 14**
1 - VGPRTRGYAI HKLGFCSVVM LGINSIIGAG IFLTPGEVIG LAGPFAPMAY VLAGIFAGVV
61 - AIVFATAARY VRTNGASYAY TTAAFGRRIG IYVGVTHAIT ASIAWGVLAS FFVSTLLRVA
121 - FPDKAWADAE QLFSVKTLTF LGFIGVLLAI NLFGNRAIKW ANGTSTVGKA FALSAFIVGG
181 - LWIITTQHVN NYATAWSAYS ATPYSLLGVA EIGKGTFSSM ALATIVALYA FTGFESIANA
241 - AEEMDAPDRN LPRAIPIAIF SVGAIYLLTL TVAMLLGSNK IAASDDTVKL AAAIGNATFR
301 - TIIVVGALIS MFGINVAASF GAPRLWTALA DSGVLPTRLS RKNQYDVPMV SFAITASLAL

Figure 7 (Cont'd)

361 - AFPLALRFDN LHLTGLAVIA RFVQFIIVPI ALIALARSQA VEHAAVRRNA FTDKVLPLVA
421 - IVVSVGLAVS YDYRCIFLVR GGPNYFSIAL IVITFVVVPA MAYLHYYRII RRVGDRPSTR

>Rv1769: 414 aa - M. tuberculosis - **SEQ. I.D. NO.** 15
1 - VHEVAAREQR SDGPMRLDAQ GRLQRYEEAF ADYDAPFAFV DLDAMWGNAD QLLARAGDKP
61 - IRVASKSLRC RPLQREILDA SERFDGLLTF TLTETLWLAG QGFSNLLLAY PPTDRAALRA
121 - LGELTAKDPD GAPIVMVDSV EHLDLIERTT DKPVRLCLDF DAGYWRAGGR IKIGSKRSPL
181 - HTPEQARALA VEIARRPALT LAALMCYEAH IAGLGDNVAG KRVHNAIIRR MQRMSFEELR
241 - ERRARAVELV REVADIKIVN AGGTGDLQLV AQEPLITEAT AGSGFYAPTL FDSYSTFTLQ
301 - PAAMFALPVC RRPGAKTVTA LGGGYLASGV GAKDRMPTPY LPVGLKLNAL EGTGEVQTPL
361 - SGDAARRLKL GDKVYFRHTK AGELCERFDH LHLVRGAEVV DTVPTYRGEG RTFL

Figure 8

>Rv1983: 1674 bp - M. tuberculosis - **SEQ. I.D. NO 64**

```
                                             ggtca
     gcgagttcggcggctagtcggtctacctcagggtctttg
     atattcagcgccacaggtagatggtaccagcaaatagcc
     actatctacctaacgcgtgctgtgccgtgcggtagctac
     tgaaaatccgagatgtcaaaggcagcgtctggatacgct
     gtatgcgcgcagggatggtgatcgaggcggagggcggc
   1 - gtg tca ttt ctg gtc gtg gtt ccc gag ttc
  31 - ttg acg tcc gcg gca gcg gat gtg gag aac
  61 - ata ggt tcc aca ctg cgc gcg gcg aat gcc
  91 - gcg gct gcc gcc tcg acc acc gcg ctt gcg
 121 - gcc gct ggc gct gat gag gta tcg gcg gcg
 151 - gtg gca gcg ctg ttt gcc agg ttc ggt cag
 181 - gaa tat caa gcg gtc agc gcg cag gcg agc
 211 - gct ttc cat caa cag ttc gtg cag acg ctg
 241 - aac tcg gcg tca gga tcg tat gcg gcc gcg
 271 - gag gcc acc atc gcg tca cag ttg cag acc
 301 - gcg cag cac gat ctg ctg ggc gcg gtc aat
 331 - gca cca acc gaa acg ttg ttg ggg cgt ccg
 361 - cta atc ggc gac gga gca ccc ggg acg gca
 391 - acg agt ccg aat ggc ggg gcg ggt ggg ctg
 421 - ctg tac ggc aac ggc ggc aac ggt tat tcc
 451 - gcg acg gcg tcg ggg gtc ggc ggc ggg gcc
 481 - ggc ggt tcc gcg ggg ttg atc ggc aat ggc
 511 - ggc gcc ggg gga gcc ggc gga ccc aac gcc
 541 - ccc ggg gga gcc ggc ggc aac ggt ggc tgg
 571 - ctg ctc ggc aac ggc ggg atc ggc ggg ccc
 601 - ggg ggc gcg tcg agc atc ccc ggc atg agt
 631 - ggt gga gcc ggc gga acc ggc ggt gcc gca
 661 - gga ctt ttg ggc tgg gga gcg aac ggc gga
 691 - gcc ggc ggc ctc ggt gat gga gtc ggt gtc
 721 - gat cgt ggc acg ggc ggc gcc gga ggc cgc
 751 - ggc ggc ctg ttg tat ggc gga tac ggc gtc
 781 - agt ggg cca ggc ggc gac ggc aga acc gtc
 811 - ccg ctg gag ata att cat gtc aca gag ccg
 841 - acg gta cat gcc aac gtc aac ggc gga ccg
 871 - acg tca acc att ctg gtc gac acc gga tcc
 901 - gct ggt ctt gtt gtc tcg cct gag gat gtc
 931 - ggg gga atc ctg gga gtg ctt cac atg ggc
 961 - ctc cca acc gga ttg agc atc agc ggt tac
 991 - agc ggg ggg ctg tac tac atc ttc gcc acg
1021- tat acc acg acg gtg gac ttc ggg aat ggc
1051- atc gtc acc gcg ccg acc gcc gtt aat gtc
1081- gtc ctc ttg tcc atc cca acg tcc ccc ttc
1111- gcc att tcg acc tac ttc agc gcc ttg ctg
1141- gcc gat ccg aca aca act ccg ttc gaa gcc
1171- tat ttc ggt gcc gtc ggc gtg gac ggc gtt
1201- ctg gga gtt ggg ccc aat gcg gtg gga cca
1231- ggc ccc agc att ccg acg atg gcg tta ccg
1261- ggt gac ctc aac cag gga gtg ctc atc gac
1291- gca ccc gca ggt gag ctc gtg ttc ggt ccc
1321- aac ccg cta cct gcg ccc aac gtc gag gtc
1351- gtc gga tcg ccg atc acc acc ctg tac gta
```

Figure 8 (Cont'd)

1381- aag atc gat ggt ggg act ccc ata ccc gtc
1411- ccc tcg atc atc gat tcc ggt ggg gta acg
1441- gga acc atc ccg tca tat gtc atc gga tcc
1471- gga acc ctg ccg gcg aac aca aac att gag
1501- gtc tac acc agc ccc ggc ggt gat cgg ctc
1531- tac gcg ttc aac aca aac gat tac cgc ccg
1561- acc gtc att tca tcc ggc ctg atg aat acc
1591- ggg ttc ttg ccc ttc aga ttc cag ccg gtg
1621- tac atc gac tac agc ccc agc ggt ata ggg
1651- aca aca gtc ttt gat cat ccg gcg
tgatcgagcctgttcgccgcgaatgtcgccgcctggctt
gtcatccccgactgaacatacgaaacatgcgccataata
ttgccgcctccggtgcatattggatcgtcgggagcacac
aagtttatggtcttagagctatacagcggaccgattgtc
ggcaacgacccgccgccccacaacatgctggagaaacca
ctgga

>Rv1983: 1674 bp - M. tuberculosis - **SEQ. I.D. NO. 11**
gtgtcatttctggtcgtggttcccgagttcttgacgtccgcggcagcggatgtggagaac
ataggttccacactgcgcgcggcgaatgccgcggctgccgcctcgaccaccgcgcttgcg
gccgctggcgctgatgaggtatcggcggcggtggcagcgctgtttgccaggttcggtcag
gaatatcaagcggtcagcgcgcaggcgagcgctttccatcaacagttcgtgcagacgctg
aactcggcgtcaggatcgtatgcggccgcggaggccaccatcgcgtcacagttgcagacc
gcgcagcacgatctgctgggcgcggtcaatgcaccaaccgaaacgttgttggggcgtccg
ctaatcggcgacggagcacccgggacggcaacgagtccgaatggcggggcgggtggcctg
ctgtacggcaacggcggcaacggttattccgcgacggcgtcggggtcggcggcggggcc
ggcggttccgcggggttgatcggcaatggcggcgccgggggagccggcggacccaacgcc
cccgggggagccggcggcaacggtggctggctgctcggcaacggcgggatcggcgggccc
gggggcgcgtcgagcatccccggcatgagtggtggagccggcggaaccggcggtgccgca
ggacttttgggctggggagcgaacggcggagccggcggcctcggtgatggagtcggtgtc
gatcgtggcacgggcggcgccggaggccgcggcggcctgttgtatggcggatacggcgtc
agtgggccaggcggcgacggcagaaccgtcccgctggagataattcatgtcacagagccg
acggtacatgccaacgtcaacggcggaccgacgtcaaccattctggtcgacaccggatcc
gctggtcttgttgtctcgcctgaggatgtcgggggaatcctgggagtgcttcacatgggc
ctcccaaccggattgagcatcagcggttacagcgggggctgtactacatcttcgccacg
tataccacgacggtggacttcgggaatggcatcgtcaccgcgccgaccgccgttaatgtc
gtcctcttgtccatcccaacgtccccttcgccatttcgacctacttcagcgccttgctg
gccgatccgacaacaactccgttcgaagcctatttcggtgccgtcggcgtggacggcgtt
ctgggagttgggcccaatgcggtgggaccaggccccagcattccgacgatggcgttaccg
ggtgacctcaaccagggagtgctcatcgacgcacccgcaggtgagctcgtgttcggtccc
aacccgctacctgcgcccaacgtcgaggtcgtcggatcgccgatcaccaccctgtacgta
aagatcgatggtgggactcccatacccgtcccctcgatcatcgattccggtggggtaacg
ggaaccatcccgtcatatgtcatcggatccggaaccctgccggcgaacacaaacattgag
gtctacaccagccccggcggtgatcggctctacgcgttcaacacaaacgattaccgcccg
accgtcatttcatccggcctgatgaataccgggttcttgcccttcagattccagccggtg
tacatcgactacagccccagcggtatagggacaacagtctttgatcatccggcg

Figure 8 (Cont'd)

>Rv1986: 597 bp - M. tuberculosis - **SEQ. I.D. NO. 61**

```
                                              tgtag
     gcgctccgcggccgcatcgaagctgcccagttcgaccac
     ggcagccaatgcggccagctgtggaccgtcaagctgcgg
     atccaccatctcaggtgtagaccatctgcggagcgtcgc
     actgcacattaataatgctaatgtaaatgaagaattatt
     agctatactgacccatacaaactgcctagtgtcgattgc
  1 - gtg aac tca cca ctg gtc gtc ggc ttc ctg
 31 - gcc tgc ttc acg ctg atc gcc gcg att ggc
 61 - gcg cag aac gca ttc gtg ctg cgg cag gga
 91 - atc cag cgt gag cac gtg ctg ccg gtg gtg
121 - gcg ctg tgc acg gtg tcc gac atc gtg ctg
151 - atc gcc gcc ggt atc gcg ggg ttc ggc gca
181 - ttg atc ggc gca cat ccg cgt gcg ctc aat
211 - gtc gtc aag ttt ggc ggc gcc gcc ttc cta
241 - atc ggc tac ggg cta ctt gcg gcc cgg cgg
271 - gcg tgg cga cct gtt gcg ctg atc cca tct
301 - ggc gcc acg ccg gtt cgc tta gcc gag gtc
331 - ctg gtg acc tgt gcg gca ttc acg ttc ctc
361 - aac cca cac gtc tac ctc gac acc gtc gtg
391 - ttg cta ggc gcg ctg gcc aac gag cac agc
421 - gac cag cgc tgg ctg ttc ggc ctc ggc gcg
451 - gtc aca gcc agt gcg gta tgg ttc gcc acc
481 - ctc ggg ttc gga gcc ggc cgg ttg cgc ggg
511 - ctg ttc acc aac ccc ggc tcg tgg aga atc
541 - ctc gac ggc ctg atc gcg gtc atg atg gtt
571 - gcg ctg gga atc tcg ctg acc gtg acc
      tagtacagcacgtgtgcacacgcgggttggaccacgtga
      tcgtcgatgggcacataccgttcggcaggagggcgcgcg
      gtcagtctgcacaactcagtcaccagctgacacgccgac
      ggcggcctcgcccggtcgtgtcggcgccaccagtgcaca
      ttcggcgtgacgcggccctacggatcgtgttggagctgt
      agccc
```

>Rv1986: 597 bp - M. tuberculosis - **SEQ. I.D. NO. 8**

```
gtgaactcaccactggtcgtcggcttcctggcctgcttcacgctgatcgccgcgattggc
gcgcagaacgcattcgtgctgcggcagggaatccagcgtgagcacgtgctgccggtggtg
gcgctgtgcacggtgtccgacatcgtgctgatcgccgccggtatcgcggggttcggcgca
ttgatcggcgcacatccgcgtgcgctcaatgtcgtcaagtttggcggcgccgccttccta
atcggctacgggctacttgcggcccggcgggcgtggcgacctgttgcgctgatcccatct
ggcgccacgccggttcgcttagccgaggtcctggtgacctgtgcggcattcacgttcctc
aacccacacgtctacctcgacaccgtcgtgttgctaggcgcgctggccaacgagcacagc
gaccagcgctggctgttcggcctcggcgcggtcacagccagtgcggtatggttcgccacc
ctcgggttcggagccggccggttgcgcgggctgttcaccaaccccggctcgtggagaatc
ctcgacggcctgatcgcggtcatgatggttgcgctgggaatctcgctgaccgtgacc
```

Figure 8 (Cont'd)

>Rv3872: 297 bp - M. tuberculosis - **SEQ. I.D. NO.** 62

```
                                     ggccc
      cctacatcgagcctccagaagaagtgttcgcagcacccc
      caagcgccggttaagattatttcattgccggtgtagcag
      gacccgagctcagcccggtaatcgagttcgggcaatgct
      gaccatcgggtttgtttccggctataaccgaacggtttg
      tgtacgggatacaaatacagggagggaagaagtaggcaa
  1 - atg gaa aaa atg tca cat gat ccg atc gct
 31 - gcc gac att ggc acg caa gtg agc gac aac
 61 - gct ctg cac ggc gtg acg gcc ggc tcg acg
 91 - gcg ctg acg tcg gtg acc ggg ctg gtt ccc
121 - gcg ggg gcc gat gag gtc tcc gcc caa gcg
151 - gcg acg gcg ttc aca tcg gag ggc atc caa
181 - ttg ctg gct tcc aat gca tcg gcc caa gac
211 - cag ctc cac cgt gcg ggc gaa gcg gtc cag
241 - gac gtc gcc cgc acc tat tcg caa atc gac
271 - gac ggc gcc gcc ggc gtc ttc gcc gaa
      taggcccccaacacatcggagggagtgatcaccatgctg
      tggcacgcaatgccaccggagctaaataccgcacggctg
      atggccggcgcgggtccggctccaatgcttgcggcggcc
      gcgggatggcagacgctttcggcggctctggacgctcag
      gccgtcgagttgaccgcgcgcctgaactctgggagaa
      gcctg
```

>Rv3872: 297 bp - M. tuberculosis - **SEQ. I.D. NO.** 9

```
atggaaaaaatgtcacatgatccgatcgctgccgacattggcacgcaagtgagcgacaac
gctctgcacggcgtgacggccggctcgacggcgctgacgtcggtgaccgggctggttccc
gcgggggccgatgaggtctccgcccaagcggcgacggcgttcacatcggagggcatccaa
ttgctggcttccaatgcatcggcccaagaccagctccaccgtgcgggcgaagcggtccag
gacgtcgcccgcacctattcgcaaatcgacgacggcgccgccggcgtcttcgccgaa
```

Figure 8 (Cont'd)

>Rv3873: 1104 bp - M. tuberculosis - **SEQ. I.D. NO 65**

```
                                                   atgag
      gtctccgcccaagcggcgacggcgttcacatcggagggc
      atccaattgctggcttccaatgcatcggcccaagaccag
      ctccaccgtgcgggcgaagcggtccaggacgtcgcccgc
      acctattcgcaaatcgacgacggcgccgccggcgtcttc
      gccgaataggcccccaacacatcggagggagtgatcacc
   1 - atg ctg tgg cac gca atg cca ccg gag cta
  31 - aat acc gca cgg ctg atg gcc ggc gcg ggt
  61 - ccg gct cca atg ctt gcg gcg gcc gcg gga
  91 - tgg cag acg ctt tcg gcg gct ctg gac gct
 121 - cag gcc gtc gag ttg acc gcg cgc ctg aac
 151 - tct ctg gga gaa gcc tgg act gga ggt ggc
 181 - agc gac aag gcg ctt gcg gct gca acg ccg
 211 - atg gtg gtc tgg cta caa acc gcg tca aca
 241 - cag gcc aag acc cgt gcg atg cag gcg acg
 271 - gcg caa gcc gcg gca tac acc cag gcc atg
 301 - gcc acg acg ccg tcg ctg ccg gag atc gcc
 331 - gcc aac cac atc acc cag gcc gtc ctt acg
 361 - gcc acc aac ttc ttc ggt atc aac acg atc
 391 - ccg atc gcg ttg acc gag atg gat tat ttc
 421 - atc cgt atg tgg aac cag gca gcc ctg gca
 451 - atg gag gtc tac cag gcc gag acc gcg gtt
 481 - aac acg ctt ttc gag aag ctc gag ccg atg
 511 - gcg tcg atc ctt gat ccc ggc gcg agc cag
 541 - agc acg acg aac ccg atc ttc gga atg ccc
 571 - tcc cct ggc agc tca aca ccg gtt ggc cag
 601 - ttg ccg ccg gcg gct acc cag acc ctc ggc
 631 - caa ctg ggt gag atg agc ggc ccg atg cag
 661 - cag ctg acc cag ccg ctg cag cag gtg acg
 691 - tcg ttg ttc agc cag gtg ggc ggc acc ggc
 721 - ggc ggc aac cca gcc gac gag gaa gcc gcg
 751 - cag atg ggc ctg ctc ggc acc agt ccg ctg
 781 - tcg aac cat ccg ctg gct ggt gga tca ggc
 811 - ccc agc gcg ggc gcg ggc ctg ctg cgc gcg
 841 - gag tcg cta cct ggc gca ggt ggg tcg ttg
 871 - acc cgc acg ccg ctg atg tct cag ctg atc
 901 - gaa aag ccg gtt gcc ccc tcg gtg atg ccg
 931 - gcg gct gct gcc gga tcg tcg gcg acg ggt
 961 - ggc gcc gct ccg gtg ggt gcg gga gcg atg
 991 - ggc cag ggt gcg caa tcc ggc ggc tcc acc
1021- agg ccg ggt ctg gtc gcg ccg gca ccg ctc
1051- gcg cag gag cgt gaa gaa gac gac gag gac
1081- gac tgg gac gaa gag gac gac tgg
        tgagctcccgtaatgacaacagacttcccggccacccgg
        gccggaagacttgccaacattttggcgaggaaggtaaag
        agagaaagtagtccagcatggcagagatgaagaccgatg
        ccgctaccctcgcgcaggaggcaggtaatttcgagcgga
        tctccggcgacctgaaaacccagatcgaccaggtggagt
        cgacg
```

Figure 8 (Cont'd)

>Rv3873: 1104 bp - M. tuberculosis - **SEQ. I.D. NO 12**

```
atgctgtggcacgcaatgccaccggagctaaataccgcacggctgatggccggcgcgggt
ccggctccaatgcttgcggcggccgcgggatggcagacgctttcggcggctctggacgct
caggccgtcgagttgaccgcgcgcctgaactctctgggagaagcctggactggaggtggc
agcgacaaggcgcttgcggctgcaacgccgatggtggtctggctacaaaccgcgtcaaca
caggccaagacccgtgcgatgcaggcgacggcgcaagccgcggcatacacccaggccatg
gccacgacgccgtcgctgccggagatcgccgccaaccacatcacccaggccgtccttacg
gccaccaacttcttcggtatcaacacgatcccgatcgcgttgaccgagatggattatttc
atccgtatgtggaaccaggcagccctggcaatggaggtctaccaggccgagaccgcggtt
aacacgcttttcgagaagctcgagccgatggcgtcgatccttgatcccggcgcgagccag
agcacgacgaacccgatcttcggaatgccctcccctggcagctcaacaccggttggccag
ttgccgccggcggctacccagaccctcggccaactgggtgagatgagcggcccgatgcag
cagctgacccagccgctgcagcaggtgacgtcgttgttcagccaggtgggcggcaccggc
ggcggcaacccagccgacgaggaagccgcgcagatgggcctgctcggcaccagtccgctg
tcgaaccatccgctggctggtggatcaggccccagcgcgggcgcgggcctgctgcgcgcg
gagtcgctacctggcgcaggtgggtcgttgacccgcacgccgctgatgtctcagctgatc
gaaaagccggttgcccccctcggtgatgccggcggctgctgccggatcgtcggcgacgggt
ggcgccgctccggtgggtgcgggagcgatgggccagggtgcgcaatccggcggctccacc
aggccgggtctggtcgcgccggcaccgctcgcgcaggagcgtgaagaagacgacgaggac
gactgggacgaagaggacgactgg
```

Figure 8 (Cont'd)

>Rv3878: 840 bp - M. tuberculosis - **SEQ. I.D. NO. 63**

tgctg
tggatcaccggggtgtacgacacggtccgcaatatccgg
ttctgagccggatcggctgattggcggttcctgacagaa
catcgaggacacggcgcaggtttgcataccttcggcgcc
cgacaaattgctgcgattgagcgtgtggcgcgtccggta
aaatttgctcgatggggaacacgtataggagatccggca
1 - atg gct gaa ccg ttg gcc gtc gat ccc acc
31 - ggc ttg agc gca gcg gcc gcg aaa ttg gcc
61 - ggc ctc gtt ttt ccg cag cct ccg gcg ccg
91 - atc gcg gtc agc gga acg gat tcg gtg gta
121 - gca gca atc aac gag acc atg cca agc atc
151 - gaa tcg ctg gtc agt gac ggg ctg ccc ggc
181 - gtg aaa gcc gcc ctg act cga aca gca tcc
211 - aac atg aac gcg gcg gcg gac gtc tat gcg
241 - aag acc gat cag tca ctg gga acc agt ttg
271 - agc cag tat gca ttc ggc tcg tcg ggc gaa
301 - ggc ctg gct ggc gtc gcc tcg gtc ggt ggt
331 - cag cca agt cag gct acc cag ctg ctg agc
361 - aca ccc gtg tca cag gtc acg acc cag ctc
391 - ggc gag acg gcc gct gag ctg gca ccc cgt
421 - gtt gtt gcg acg gtg ccg caa ctc gtt cag
451 - ctg gct ccg cac gcc gtt cag atg tcg caa
481 - aac gca tcc ccc atc gct cag acg atc agt
511 - caa acc gcc caa cag gcc gcc cag agc gcg
541 - cag ggc ggc agc ggc cca atg ccc gca cag
571 - ctt gcc agc gct gaa aaa ccg gcc acc gag
601 - caa gcg gag ccg gtc cac gaa gtg aca aac
631 - gac gat cag ggc gac cag ggc gac gtg cag
661 - ccg gcc gag gtc gtt gcc gcg gca cgt gac
691 - gaa ggc gcc ggc gca tca ccg ggc cag cag
721 - ccc ggc ggg ggc gtt ccc gcg caa gcc atg
751 - gat acc gga gcc ggt gcc cgc cca gcg gcg
781 - agt ccg ctg gcg gcc ccc gtc gat ccg tcg
811 - act ccg gca ccc tca aca acc aca acg ttg

tagaccgggcctgccagcggctccgtctcgcacgcagcg
cctgttgctgtcctggcctcgtcagcatgcggcggccag
ggcccggtcgagcaacccggtgacgtattgccagtacag
ccagtccgcgacggccacacgctggacggccgcgtcagt
cgcagtgtgcgcttggtgcagggcaatctcctgtgagtg
ggcag

>Rv3878: 840 bp - M. tuberculosis - **SEQ. I.D. NO. 10**
atggctgaaccgttggccgtcgatcccaccggcttgagcgcagcggccgcgaaattggcc
ggcctcgtttttccgcagcctccggcgccgatcgcggtcagcggaacggattcggtggta
gcagcaatcaacgagaccatgccaagcatcgaatcgctggtcagtgacgggctgcccggc
gtgaaagccgccctgactcgaacagcatccaacatgaacgcggcggcggacgtctatgcg
aagaccgatcagtcactgggaaccagtttgagccagtatgcattcggctcgtcgggcgaa
ggcctggctggcgtcgcctcggtcggtggtcagccaagtcaggctacccagctgctgagc
acacccgtgtcacaggtcacgacccagctcggcgagacggccgctgagctggcaccccgt
gttgttgcgacggtgccgcaactcgttcagctggctccgcacgccgttcagatgtcgcaa
aacgcatcccccatcgctcagacgatcagtcaaaccgcccaacaggccgcccagagcgcg

Figure 8 (Cont'd)

cagggcggcagcggcccaatgcccgcacagcttgccagcgctgaaaaccggccaccgag
caagcggagccggtccacgaagtgacaaacgacgatcagggcgaccagggcgacgtgcag
ccggccgaggtcgttgccgcggcacgtgacgaaggcgccggcgcatcaccgggccagcag
cccggcgggggcgttcccgcgcaagccatggataccggagccggtgcccgcccagcggcg
agtccgctggcggcccccgtcgatccgtcgactccggcaccctcaacaaccacaacgttg

Figure 8 (Cont'd)

>Rv3879c: 2187 bp - M. tuberculosis - **SEQ. I.D. NO. 66**

cccgt
gcgacagcgcggccagcgcaagcgaggtgaccacccggc
tgatcgcccaagcggtgcatcatgcgcgcggattcaacg
ggttactgcgaataccggcgcggggtggatccagcggcc
gagccggcgtgaaatgccggaggccaaccggacggtgat
ccgcgaggcgatctggcggtttggggagggcagtagggg

1 - atg agt att acc agg ccg acg ggc agc tat
31 - gcc aga cag atg ctg gat ccg ggc ggc tgg
61 - gtg gaa gcc gat gaa gac act ttc tat gac
91 - cgg gcc cag gaa tat agc cag gtt ttg caa
121 - agg gtc acc gat gta ttg gac acc tgc cgc
151 - cag cag aaa ggc cac gtc ttc gaa ggc ggc
181 - cta tgg tcc ggc ggc gcc gcc aat gct gcc
211 - aac ggc gcc ctg ggt gca aac atc aat caa
241 - ttg atg acg ctg cag gat tat ctc gcc acg
271 - gtg att acc tgg cac agg cat att gcc ggg
301 - ttg att gag caa gct aaa tcc gat atc ggc
331 - aat aat gtg gat ggc gct caa cgg gag atc
361 - gat atc ctg gag aat gac cct agc ctg gat
391 - gct gat gag cgc cat acc gcc atc aat tca
421 - ttg gtc acg gcg acg cat ggg gcc aat gtc
451 - agt ctg gtc gcc gag acc gct gag cgg gtg
481 - ctg gaa tcc aag aat tgg aaa cct ccg aag
511 - aac gca ctc gag gat ttg ctt cag cag aag
541 - tcg ccg cca ccc cca gac gtg cct acc ctg
571 - gtc gtg cca tcc ccg ggc aca ccg ggc aca
601 - ccg gga acc ccg atc acc ccg gga acc ccg
631 - atc acc ccg gga acc cca atc aca ccc atc
661 - ccg gga gcg ccg gta act ccg atc aca cca
691 - acg ccc ggc act ccc gtc acg ccg gtg acc
721 - ccg ggc aag ccg gtc acc ccg gtg acc ccg
751 - gtc aaa ccg ggc aca cca ggc gag cca acc
781 - ccg atc acg ccg gtc acc ccc ccg gtc gcc
811 - ccg gcc aca ccg gca acc ccg gcc acg ccc
841 - gtt acc cca gct ccc gct cca cac ccg cag
871 - ccg gct ccg gca ccg gcg cca tcg cct ggg
901 - ccc cag ccg gtt aca ccg gcc act ccc ggt
931 - ccg tct ggt cca gca aca ccg ggc acc cca
961 - ggg ggc gag ccg gcg ccg cac gtc aaa ccc
991 - gcg gcg ttg gcg gag caa cct ggt gtg ccg
1021- ggc cag cat gcg ggc ggg ggg acg cag tcg
1051- ggg cct gcc cat gcg gac gaa tcc gcc gcg
1081- tcg gtg acg ccg gct gcg gcg tcc ggt gtc
1111- ccg ggc gca cgg gcg gcg gcc gcc gcg ccg
1141- agc ggt acc gcc gtg gga gcg ggc gcg cgt
1171- tcg agc gtg ggt acg gcc gcg gcc tcg ggc
1201- gcg ggg tcg cat gct gcc act ggg cgg gcg
1231- ccg gtg gct acc tcg gac aag gcg gcg gca
1261- ccg agc acg cgg gcg gcc tcg gcg cgg acg
1291- gca cct cct gcc cgc ccg ccg tcg acc gat
1321- cac atc gac aaa ccc gat cgc agc gag tct
1351- gca gat gac ggt acg ccg gtg tcg atg atc

Figure 8 (Cont'd)

1381- ccg gtg tcg gcg gct cgg gcg gca cgc gac
1411- gcc gcc act gca gct gcc agc gcc cgc cag
1441- cgt ggc cgc ggt gat gcg ctg cgg ttg gcg
1471- cga cgc atc gcg gcg gcg ctc aac gcg tcc
1501- gac aac aac gcg ggc gac tac ggg ttc ttc
1531- tgg atc acc gcg gtg acc acc gac ggt tcc
1561- atc gtc gtg gcc aac agc tat ggg ctg gcc
1591- tac ata ccc gac ggg atg gaa ttg ccg aat
1621- aag gtg tac ttg gcc agc gcg gat cac gca
1651- atc ccg gtt gac gaa att gca cgc tgt gcc
1681- acc tac ccg gtt ttg gcc gtg caa gcc tgg
1711- gcg gct ttc cac gac atg acg ctg cgg gcg
1741- gtg atc ggt acc gcg gag cag ttg gcc agt
1771- tcg gat ccc ggt gtg gcc aag att gtg ctg
1801- gag cca gat gac att ccg gag agc ggc aaa
1831- atg acg ggc cgg tcg cgg ctg gag gtc gtc
1861- gac ccc tcg gcg gcg gct cag ctg gcc gac
1891- act acc gat cag cgt ttg ctc gac ttg ttg
1921- ccg ccg gcg ccg gtg gat gtc aat cca ccg
1951- ggc gat gag cgg cac atg ctg tgg ttc gag
1981- ctg atg aag ccc atg acc agc acc gct acc
2011- ggc cgc gag gcc gct cat ctg cgg gcg ttc
2041- cgg gcc tac gct gcc cac tca cag gag att
2071- gcc ctg cac caa gcg cac act gcg act gac
2101- gcg gcc gtc cag cgt gtg gcc gtc gcg gac
2131- tgg ctg tac tgg caa tac gtc acc ggg ttg
2161- ctc gac cgg gcc ctg gcc gcc gca tgc
tgacgaggccaggacagcaacaggcgctgcgtgcgagac
ggagccgctggcaggcccggtctacaacgttgtggttgt
tgagggtgccggagtcgacggatcgacgggggccgccag
cggactcgccgctgggcgggcaccggctccggtatccat
ggcttgcgcgggaacgcccccgccgggctgctggcccgg
tgatg

>Rv3879c: 2187 bp - M. tuberculosis - **SEQ. I.D. NO. 13**
atgagtattaccaggccgacgggcagctatgccagacagatgctggatccgggcggctgg
gtggaagccgatgaagacactttctatgaccgggcccaggaatatagccaggttttgcaa
agggtcaccgatgtattggacacctgccgccagcagaaaggccacgtcttcgaaggcggc
ctatggtccggcggcgccgccaatgctgccaacggcgccctgggtgcaaacatcaatcaa
ttgatgacgctgcaggattatctcgccacggtgattacctggcacaggcatattgccggg
ttgattgagcaagctaaatccgatatcggcaataatgtggatggcgctcaacgggagatc
gatatcctggagaatgaccctagcctggatgctgatgagcgccataccgccatcaattca
ttggtcacggcgacgcatggggccaatgtcagtctggtcgccgagaccgctgagcgggtg
ctggaatccaagaattggaaacctccgaagaacgcactcgaggatttgcttcagcagaag
tcgccgccacccccagacgtgcctaccctggtcgtgccatccccgggcacaccgggcaca
ccgggaaccccgatcaccccgggaaccccgatcaccccgggaaccccaatcacacccatc
ccgggagcgccggtaactccgatcacaccaacgcccggcactcccgtcacgccggtgacc
ccgggcaagccggtcaccccggtgaccccggtcaaaccgggcacaccaggcgagccaacc
ccgatcacgccggtcacccccccggtcgccccggccacaccggcaaccccggccacgccc
gttaccccagctcccgctccacacccgcagccggctccggcaccggcgccatcgcctggg
ccccagccggttacaccggccactcccggtccgtctggtccagcaacaccgggcacccca
gggggcgagccggcgccgcacgtcaaacccgcggcgttggcggagcaacctggtgtgccg
ggccagcatgcgggcgggggacgcagtcggggcctgcccatgcggacgaatccgccgcg

Figure 8 (Cont'd)

tcggtgacgccggctgcggcgtccggtgtcccgggcgcacgggcggcggccgccgcgccg
agcggtaccgccgtgggagcgggcgcgcgttcgagcgtgggtacggccgcggcctcgggc
gcggggtcgcatgctgccactgggcgggcgccggtggctacctcggacaaggcggcggca
ccgagcacgcgggcggcctcggcgcggacggcacctcctgcccgcccgccgtcgaccgat
cacatcgacaaacccgatcgcagcgagtctgcagatgacggtacgccggtgtcgatgatc
ccggtgtcggcggctcgggcggcacgcgacgccgccactgcagctgccagcgcccgccag
cgtggccgcggtgatgcgctgcggttggcgcgacgcatcgcggcggcgctcaacgcgtcc
gacaacaacgcgggcgactacgggttcttctggatcaccgcggtgaccaccgacggttcc
atcgtcgtggccaacagctatgggctggcctacatacccgacgggatggaattgccgaat
aaggtgtacttggccagcgcggatcacgcaatcccggttgacgaaattgcacgctgtgcc
acctacccggttttggccgtgcaagcctgggcggctttccacgacatgacgctgcgggcg
gtgatcggtaccgcggagcagttggccagttcggatcccggtgtggccaagattgtgctg
gagccagatgacattccggagagcggcaaaatgacgggccggtcgcggctggaggtcgtc
gacccctcggcggcggctcagctggccgacactaccgatcagcgtttgctcgacttgttg
ccgccggcgccggtggatgtcaatccaccgggcgatgagcggcacatgctgtggttcgag
ctgatgaagcccatgaccagcaccgctaccggccgcgaggccgctcatctgcgggcgttc
cgggcctacgctgcccactcacaggagattgccctgcaccaagcgcacactgcgactgac
gcggccgtccagcgtgtggccgtcgcggactggctgtactggcaatacgtcaccgggttg
ctcgaccgggccctggccgccgcatgc

Figure 9

Diagnostic Cocktail 1

SEQ ID NO 23;

AT AQAAAYTQAM ATTPSLPEIA ANHITQAVLT ATNFFGINTI PIALTEMDYF IRMWNQAALA
MEVYQAETAV NTLFEKLEPM ASILDPGASQ STTNPIFG

Derived from the sequence of a segment of Rv3873 (SEQ ID NO 5), as highlighted in bold below -

>M. tuberculis bacteria|Rv3873|PPE68: 368 aa - PPE FAMILY PROTEIN

1 - MLWHAMPPEL NTARLMAGAG PAPMLAAAAG WQTLSAALDA QAVELTARLN SLGEAWTGGG
61 - SDKALAAATP MVVWLQTAST QAKTRAMQAT **AQAAAYTQAM ATTPSLPEIA ANHITQAVLT**
121 - **ATNFFGINTI PIALTEMDYF IRMWNQAALA MEVYQAETAV NTLFEKLEPM ASILDPGASQ**
181 - **STTNPIFGMP** SPGSSTPVGQ LPPAATQTLG QLGEMSGPMQ QLTQPLQQVT SLFSQVGGTG
241 - GGNPADEEAA QMGLLGTSPL SNHPLAGGSG PSAGAGLLRA ESLPGAGGSL TRTPLMSQLI
301 - EKPVAPSVMP AAAAGSSATG GAAPVGAGAM GQGAQSGGST RPGLVAPAPL AQEREEDDED
361 - DWDEEDDW

Cocktail comprised of 11 peptides, each 20 amino acids long, with an overlap of 12 residues.

ATAQAAAYTQAMATTPSLPE	SEQ ID NO 24
TQAMATTPSLPEIAANHITQ	SEQ ID NO 25
SLPEIAANHITQAVLTATNF	SEQ ID NO 26
HITQAVLTATNFFGINTIPI	SEQ ID NO 27
ATNFFGINTIPIALTEMDYF	**SEQ ID NO 28**
TIPIALTEMDYFIRMWNQAA	**SEQ ID NO 29**
MDYFIRMWNQAALAMEVYQA	SEQ ID NO 30
NQAALAMEVYQAETAVNTLF	SEQ ID NO 31
VYQAETAVNTLFEKLEPMAS	SEQ ID NO 32
NTLFEKLEPMASILDPGASQ	SEQ ID NO 33
PMASILDPGASQSTTNPIFG	SEQ ID NO 34

Peptides highlighted in bold are of special importance, data suggesting the main epitope of the pool lies within.

Figure 10

Diagnostic Cocktail 2

SEQ ID NO 35:

AQSA QGGSGPMPAQ LASAEKPATE QAEPVHEVTN DDQGDQGDVQ PAEVVAAARD EGAGASPGQQ
PGGGVPAQAM DTGAGARPAA SPLAAPVDPS TPAPSTTTTL

Derived from the sequence of a segment of Rv3878 (SEQ ID NO 3), highlighted in bold below -

>M. tuberculis bacteria|Rv3878|Rv3878: 280 aa - CONSERVED HYPOTHETICAL ALANINE RICH PROTEIN

1 - MAEPLAVDPT GLSAAAAKLA GLVFPQPPAP IAVSGTDSVV AAINETMPSI ESLVSDGLPG
61 - VKAALTRTAS NMNAAADVYA KTDQSLGTSL SQYAFGSSGE GLAGVASVGG QPSQATQLLS
121 - TPVSQVTTQL GETAAELAPR VVATVPQLVQ LAPHAVQMSQ NASPIAQTIS QTAQQ**AAQSA**
181 - **QGGSGPMPAQ LASAEKPATE QAEPVHEVTN DDQGDQGDVQ PAEVVAAARD EGAGASPGQQ**
241 - **PGGGVPAQAM DTGAGARPAA SPLAAPVDPS** TPAPSTTTTL

Cocktail comprised of 12 peptides, each 20 amino acids long, with an overlap of 12 residues.

AQSAQGGSGPMPAQLASAEK	SEQ ID NO 36
GPMPAQLASAEKPATEQAEP	SEQ ID NO 37
SAEKPATEQAEPVHEVTNDD	SEQ ID NO 38
QAEPVHEVTNDDQGDQGDVQ	SEQ ID NO 39
TNDDQGDQGDVQPAEVVAAA	SEQ ID NO 40
GDVQPAEVVAAARDEGAGAS	SEQ ID NO 41
VAAARDEGAGASPGQQPGGG	SEQ ID NO 42
AGASPGQQPGGGVPAQAMDT	SEQ ID NO 43
PGGGVPAQAMDTGAGARPAA	SEQ ID NO 44
AMDTGAGARPAASPLAAPVD	SEQ ID NO 45
RPAASPLAAPVDPSTPAPST	SEQ ID NO 46
SPLAAPVDPSTPAPSTTTTL	SEQ ID NO 47

Figure 11

Diagnostic Cocktail 3

SEQ ID NO 48

MSITRPTGSY ARQMLDPGGW VEADEDTFYD RAQEYSQVLQ RVTDVLDTCR QQKGHVFEGG
LWSGGAANAA NGALGANINQ LMTLQDYLAT VI

Derived from the sequence of a segment of Rv3879c (SEQ ID NO 6), highlighted in bold below –

>M. tuberculis bacteria|Rv3879c|Rv3879c: 729 aa - HYPOTHETICAL ALANINE AND PROLINE RICH PROTEIN

1 – **MSITRPTGSY ARQMLDPGGW VEADEDTFYD RAQEYSQVLQ RVTDVLDTCR QQKGHVFEGG**
61 – **LWSGGAANAA NGALGANINQ LMTLQDYLAT VI**TWHRHIAG LIEQAKSDIG NNVDGAQREI
121 – DILENDPSLD ADERHTAINS LVTATHGANV SLVAETAERV LESKNWKPPK NALEDLLQQK
181 – SPPPPDVPTL VVPSPGTPGT PGTPITPGTP ITPGTPITPI PGAPVTPITP TPGTPVTPVT
241 – PGKPVTPVTP VKPGTPGEPT PITPVTPPVA PATPATPATP VTPAPAPHPQ PAPAPAPSPG
301 – PQPVTPATPG PSGPATPGTP GGEPAPHVKP AALAEQPGVP GQHAGGGTQS GPAHADESAA
361 – SVTPAAASGV PGARAAAAAP SGTAVGAGAR SSVGTAAASG AGSHAATGRA PVATSDKAAA
421 – PSTRAASART APPARPPSTD HIDKPDRSES ADDGTPVSMI PVSAARAARD AATAAASARQ
481 – RGRGDALRLA RRIAAALNAS DNNAGDYGFF WITAVTTDGS IVVANSYGLA YIPDGMELPN
541 – KVYLASADHA IPVDEIARCA TYPVLAVQAW AAFHDMTLRA VIGTAEQLAS SDPGVAKIVL
601 – EPDDIPESGK MTGRSRLEVV DPSAAAQLAD TTDQRLLDLL PPAPVDVNPP GDERHMLWFE
661 – LMKPMTSTAT GREAAHLRAF RAYAAHSQEI ALHQAHTATD AAVQRVAVAD WLYWQYVTGL
721 – LDRALAAAC

Cocktail comprised of 10 peptides, each 20 amino acids long, with an overlap of 12 residues.

MSITRPTGSYARQMLDPGGW	SEQ ID NO 49
SYARQMLDPGGWVEADEDTF	SEQ ID NO 50
PGGWVEADEDTFYDRAQEYS	**SEQ ID NO 51**
EDTFYDRAQEYSQVLQRVTD	SEQ ID NO 52
QEYSQVLQRVTDVLDTCRQQ	SEQ ID NO 53
RVTDVLDTCRQQKGHVFEGG	SEQ ID NO 54
CRQQKGHVFEGGLWSGGAAN	SEQ ID NO 55
FEGGLWSGGAANAANGALGA	SEQ ID NO 56
GAANAANGALGANINQLMTL	SEQ ID NO 57
ALGANINQLMTLQDYLATVI	SEQ ID NO 58

Peptide highlighted in bold is of special importance, data suggesting the main epitope of the pool lies within it.

Figure 12

>Rv1979c: 1443 bp - M. tuberculosis - SEQ. I.D. NO. 70

```
                                        cgact
     cgatgctggcctagactcgcgaggaccgcgcggtggtca
     ctgcgcggatttggggcggcggaaatgagtgttcggtgc
     gcccactgcggtgactcacctgcagcgccggcatcgaca
     ggccgggagctcaagaatcgtcgctagagaatctatggt
     gcgttagaggattccctgctagacagccttggtgcggtg
   1 - gtc ggc ccg cgg acg aga gga tat gcg atc
  31 - cac aag ctg ggt ttc tgc agc gtc gtc atg
  61 - ctc ggg atc aac tcg ata atc ggc gcc ggt
  91 - atc ttc cta act cca ggt gag gtg atc ggg
 121 - ctc gca gga ccc ttc gcg ccg atg gcc tat
 151 - gtt tta gct ggc att ttc gcg ggt gtc gtg
 181 - gcg atc gtc ttc gcg acg gcg gca agg tac
 211 - gtc aga aca aac ggt gcc tcc tac gcc tac
 241 - aca acg gcc gca ttt ggg cgc cgg atc ggc
 271 - atc tat gtc ggt gtc acc cac gcc att acc
 301 - gcg tcc atc gct tgg ggg gtg ttg gct tct
 331 - ttt ttc gtc tcg acg ctg ttg cga gtg gcc
 361 - ttc ccc gac aag gcc tgg gcc gac gcc gag
 391 - caa ctg ttc agt gtg aag acg ctg acg ttt
 421 - ctc ggc ttt atc ggc gtg ctg ttg gcc atc
 451 - aac ctc ttc ggc aac cgg gcg atc aag tgg
 481 - gcc aac gga acg tca acg gta ggc aag gca
 511 - ttc gcg ctc tcg gca ttc att gtc ggc ggg
 541 - ctg tgg atc atc acc acc cag cac gtg aac
 571 - aac tac gca acg gcg tgg tcg gca tac agc
 601 - gcg acc ccg tac tcg ttg ctt ggc gtc gcc
 631 - gaa att ggc aag ggc acg ttc tcg agt atg
 661 - gcg ctg gcc acg att gtc gcg ttg tac gca
 691 - ttc acc ggt ttc gaa tcg atc gcg aac gcc
 721 - gcc gaa gaa atg gac gcg ccg gac cgg aac
 751 - ctg ccg aga gct ata ccg atc gcg atc ttc
 781 - tcg gtt ggc gcg atc tac ttg ctc acc cta
 811 - acg gta gcg atg ctg ctc gga tcg aac aag
 841 - atc gcc gcg tcg gac gac acc gtg aaa ctg
 871 - gcc gcg gcc atc gga aac gct acc ttc cga
 901 - acg atc atc gtc gtc gga gcc ctg ata tcg
 931 - atg ttc ggc atc aat gtc gcg gcc tcg ttc
 961 - ggt gca ccg cgg ctt tgg acc gcg tta gcg
 991 - gac agc ggg gtt ctg ccg aca cgc ttg tca
1021- cgc aag aac caa tac gac gtg ccg atg gtc
1051- tcc ttc gca att acg gcg tcg ttg gcg ctc
1081- gca ttc ccg ttg gcg ctg cgg ttc gac aac
1111- ctg cac ctg acc ggc ctg gcg gtg atc gcc
1141- cga ttc gtc cag ttc atc atc gtg ccg atc
1171- gct ctc atc gca ttg gcg agg tct cag gca
1201- gta gaa cat gct gct gtg cgg cga aat gcg
1231- ttc acc gac aag gtg tta ccg ctt gtt gcg
1261- atc gtg gtc tcg gtt ggg ctg gca gtg tcc
1291- tac gac tac cgc tgc atc ttt cta gtg cgg
1321- ggt ggt ccg aac tac ttc tcg att gct ttg
```

Figure 12 (Cont'd)

```
1351- atc gtg atc acg ttc gtc gtg gta ccg gcg
1381- atg gct tat ctg cac tac tac cga atc att
1411- cgc cgg gtt ggc gat cgg ccg agc act cgc
1441- tag
      attccgttggcgctgagctcgaacgggagaacacaacgg
      cgagcgatggcgggaatagcctggtcggtgcgggcaaga
      tttcaacctgcattcccggatcggcggcgcgggcaagcg
      tctgcaacgccgagggactgtaggcacgtagtgcgctga
      taaagccgtcgtgcatgctcgagcgcatcgacgaccatg
      gcagc
```

>Rv1979c: 1443 bp - M. tuberculosis - **SEQ. I.D. NO. 59**

```
gtcggcccgcggacgagaggatatgcgatccacaagctgggtttctgcagcgtcgtcatg
ctcgggatcaactcgataatcggcgccggtatcttcctaactccaggtgaggtgatcggg
ctcgcaggacccttcgcgccgatggcctatgttttagctggcattttcgcgggtgtcgtg
gcgatcgtcttcgcgacggcggcaaggtacgtcagaacaaacggtgcctcctacgcctac
acaacggccgcatttgggcgccggatcggcatctatgtcggtgtcacccacgccattacc
gcgtccatcgcttgggggtgttggcttctttttcgtctcgacgctgttgcgagtggcc
ttccccgacaaggcctgggccgacgccgagcaactgttcagtgtgaagacgctgacgttt
ctcggctttatcggcgtgctgttggccatcaacctcttcggcaaccgggcgatcaagtgg
gccaacggaacgtcaacggtaggcaaggcattcgcgctctcggcattcattgtcggcggg
ctgtggatcatcaccacccagcacgtgaacaactacgcaacggcgtggtcggcatacagc
gcgaccccgtactcgttgcttggcgtcgccgaaattggcaagggcacgttctcgagtatg
gcgctggccacgattgtcgcgttgtacgcattcaccggtttcgaatcgatcgcgaacgcc
gccgaagaaatggacgcgccggaccggaacctgccgagagctataccgatcgcgatcttc
tcggttggcgcgatctacttgctcaccctaacggtagcgatgctgctcggatcgaacaag
atcgccgcgtcggacgacaccgtgaaactggccgcggccatcggaaacgctaccttccga
acgatcatcgtcgtcggagccctgatatcgatgttcggcatcaatgtcgcggcctcgttc
ggtgcaccgcggctttggaccgcgttagcggacagcggggttctgccgacacgcttgtca
cgcaagaaccaatacgacgtgccgatggtctccttcgcaattacggcgtcgttggcgctc
gcattcccgttggcgctgcggttcgacaacctgcacctgaccggcctggcggtgatcgcc
cgattcgtccagttcatcatcgtgccgatcgctctcatcgcattggcgaggtctcaggca
gtagaacatgctgctgtgcggcgaaatgcgttcaccgacaaggtgttaccgcttgttgcg
atcgtggtctcggttgggctggcagtgtcctacgactaccgctgcatctttctagtgcgg
ggtggtccgaactacttctcgattgctttgatcgtgatcacgttcgtcgtggtaccggcg
atggcttatctgcactactaccgaatcattcgccgggttggcgatcggccgagcactcgc
tag
```

Figure 12 (Cont'd)

>Rv1769: 1242 bp - M. tuberculosis - **SEQ. I.D. NO.71**

```
                                          tcggg
      cgggttgctattcggccaaaatgggatgcccgggccgtg
      agcgccccaacccaggccaacccccctatgggcaatctgc
      acatcaattggccaggtcgacagcagaccgcacacatct
      acgagattggttcccgatccgtgggtggggccgggaaaa
      gcggctgtaagagttggctaggttcagtagggtggcggc
   1 - gtg cat gag gtg gct gct cgt gag caa cgt
  31 - tcg gac ggg ccg atg agg ctg gat gcg cag
  61 - ggc cga ctg cag cgt tac gag gag gcg ttc
  91 - gct gac tac gat gca ccg ttt gcg ttc gta
 121 - gat ctc gac gcg atg tgg ggc aat gcc gat
 151 - caa ctg ctt gcg cgc gcc ggc gac aag ccg
 181 - atc cgg gtg gcg tcg aag tcg ctg cgt tgc
 211 - cga cca ctg caa cgc gaa atc ctt gat gcc
 241 - agt gag cga ttc gac ggg cta ttg acg ttc
 271 - acg ctt acc gag acg ctg tgg ctt gcc ggc
 301 - caa ggt ttc tcg aac ctg ttg ttg gcc tac
 331 - ccg ccg acc gac cgg gcg gca ttg cgt gcg
 361 - ctt ggc gag ctg acg gcc aag gac ccg gac
 391 - ggg gcg ccg atc gtg atg gtg gac agc gtg
 421 - gag cac ctt gac ctg atc gag cgc acg acc
 451 - gac aag ccg gta cgg ctg tgt ctg gat ttc
 481 - gat gcc ggc tat tgg cgc gcc ggc ggg cgg
 511 - ata aaa att ggt tcc aag cgc tcg ccg ctg
 541 - cac acc ccg gag cag gct cgc gca ctc gcg
 571 - gtg gag atc gcg cgg cgg ccg gcg cta acg
 601 - ttg gcg gcg ttg atg tgc tac gag gcc cac
 631 - att gcg ggc ctc ggt gac aac gtc gcc ggc
 661 - aag cgg gtc cac aac gcg atc atc cgt cgg
 691 - atg cag cgc atg tcg ttc gaa gag ctg cgc
 721 - gag cgt cgt gcc cgg gcc gtc gag ctg gtg
 751 - cgc gag gtc gcc gac atc aag atc gtc aac
 781 - gcc ggt ggc acc ggc gac ttg cag ctg gtt
 811 - gcg cag gag ccg ttg att acc gaa gcg acc
 841 - gcc ggc tcg ggt ttt tac gcg ccg aca ctg
 871 - ttc gac tcg tat tcg acg ttc acg ctg cag
 901 - ccc gcg gcg atg ttc gcg ctg ccg gta tgc
 931 - cgt cgt ccc ggt gca aag acc gtg acc gcg
 961 - ctc ggg ggt ggc tat tta gcc agc ggg gtc
 991 - ggg gcg aag gac cgc atg ccg act ccc tac
1021- ctg ccg gtc ggg ctg aag ctc aat gcg ctg
1051- gag gga acg ggc gaa gtt cag aca ccg cta
1081- tcc ggt gat gca gcc cga cgg ctg aag ctt
1111- ggc gac aag gtc tac ttc cgc cac acc aag
1141- gcc ggt gag ctg tgt gag cgg ttc gac cat
1171- ctg cat ctg gtc cgt ggc gct gaa gta gtc
1201- gac acc gtc ccc acc tac cgg ggt gaa ggg
1231- cgc acc ttc ctc
       taatgctgaaatggacgaggcccacccggctcacccggc
       agatgcggggcggcccggtggcccaattcaaggcgcgcg
       aagaggagctgccatgacaccgatcaccgccctgccgac
```

Figure 12 (Cont'd)

cgagttggcggccatgcgcgaggtagtcgagacgctcgc
acccattgagcgtgccgcgggcgagccgggtgagcacaa
ggcgg

>Rv1769: 1242 bp - M. tuberculosis - **SEQ. I.D. NO.60**
gtgcatgaggtggctgctcgtgagcaacgttcggacgggccgatgaggctggatgcgcag
ggccgactgcagcgttacgaggaggcgttcgctgactacgatgcaccgtttgcgttcgta
gatctcgacgcgatgtggggcaatgccgatcaactgcttgcgcgcgccggcgacaagccg
atccgggtggcgtcgaagtcgctgcgttgccgaccactgcaacgcgaaatccttgatgcc
agtgagcgattcgacgggctattgacgttcacgcttaccgagacgctgtggcttgccggc
caaggtttctcgaacctgttgttggcctacccgccgaccgaccgggcggcattgcgtgcg
cttggcgagctgacggccaaggacccggacggggcgccgatcgtgatggtggacagcgtg
gagcaccttgacctgatcgagcgcacgaccgacaagccggtacggctgtgtctggatttc
gatgccggctattggcgcgccggcgggcggataaaaattggttccaagcgctcgccgctg
cacaccccggagcaggctcgcgcactcgcggtggagatcgcgcggcggccggcgctaacg
ttggcggcgttgatgtgctacgaggcccacattgcgggcctcggtgacaacgtcgccggc
aagcgggtccacaacgcgatcatccgtcggatgcagcgcatgtcgttcgaagagctgcgc
gagcgtcgtgcccgggccgtcgagctggtgcgcgaggtcgccgacatcaagatcgtcaac
gccggtggcaccggcgacttgcagctggttgcgcaggagccgttgattaccgaagcgacc
gccggctcgggtttttacgcgccgacactgttcgactcgtattcgacgttcacgctgcag
cccgcggcgatgttcgcgctgccggtatgccgtcgtcccggtgcaaagaccgtgaccgcg
ctcggggggtggctatttagccagcggggtcggggcgaaggaccgcatgccgactccctac
ctgccggtcgggctgaagctcaatgcgctggagggaacgggcgaagttcagacaccgcta
tccggtgatgcagcccgacggctgaagcttggcgacaaggtctacttccgccacaccaag
gccggtgagctgtgtgagcggttcgaccatctgcatctggtccgtggcgctgaagtagtc
gacaccgtccccacctaccggggtgaagggcgcaccttcctc

Figure 13

Complete Vaccine Sequence

Highlighted in bold the position and sequence of Fusion Insert

```
   1  aatgacggta aatggcccgc ctggcattat gcccagtaca tgaccttatg
       q  -  r   -  m  a  r   l  a  l   c  p  v   h  d  l  m
  51  ggactttcct acttggcagt acatctacgt attagtcatc gctattacca
        g  l  s   y  l  a   v  h  l  r   i  s  h   r  y  y
 101  tggtgatgcg gttttggcag tacatcaatg ggcgtggata gcggtttgac
       h  g  d  a   v  l  a   v  h  q   w  a  w  i   a  v  -
 151  tcacggggat ttccaagtct ccaccccatt gacgtcaatg ggagtttgtt
        l  t  g   i  s  k  s   p  p  h   -  r  q   w  e  f  v
 201  ttggcaccaa aatcaacggg actttccaaa atgtcgtaac aactccgccc
        l  a  p   k  s  t   g  l  s  k   m  s  -   q  l  r
 251  cattgacgca aatgggcggt aggcgtgtac ggtgggaggt ctatataagc
       p  i  d  a   n  g  r   -  a  c   t  v  g  g   l  y  k
 301  agagctctct ggctaactag agaacccact gcttactggc ttatcgaaat
        q  s  s   l  a  n  -   r  t  h   c  l  l   a  y  r  n
 351  taatacgact cactataggg agacccaagc ttagacgcct ggagacgcca
        -  y  d   s  l  -   g  d  p  s   l  d  a   w  r  r
 401  tccacgctgt tttgacctcc atagaagaca ccgggaccga tccagcctcc
       h  p  r  c   f  d  l   h  r  r   h  r  d  r   s  s  l
 451  gcggccggga acggtgcatt ggaacgcgga ttccccgtgc caagagtgac
        r  g  r   e  r  c  i   g  t  r   i  p  r   a  k  s  d
 501  gtaagtaccg cctatagagt ctataggccc accccttgg cttcttatgc
        v  s  t   a  y  r   v  y  r  p   t  p  l   a  s  y
 551  atgctatact gtttttggct tggggtctat acacccccgc ttcctcatgt
       a  c  y  t   v  f  g   l  g  s   i  h  p  r   f  l  m
 601  tataggtgat ggtatagctt agcctatagg tgtgggttat tgaccattat
        l  -  v   m  v  -  l   s  l  -   v  w  v   i  d  h  y
 651  tgaccactcc cctattggtg acgatacttt ccattactaa tccataacat
        -  p  l   p  y  w   -  r  y  f   p  l  l   i  h  n
 701  ggctctttgc cacaactctc tttattggct atatgccaat acactgtcct
       m  a  l  c   h  n  s   l  y  w   l  y  a  n   t  l  s
 751  tcagagactg acacggactc tgtattttta caggatgggg tctcatttat
        f  r  d   -  h  g  l   c  i  f   t  g  w   g  l  i  y
 801  tatttacaaa ttcacatata caacaccacc gtccccagtg cccgcagttt
        y  l  q   i  h  i   y  n  t  t   v  p  s   a  r  s
 851  ttattaaaca taacgtggga tctccacgcg aatctcgggt acgtgttccg
       f  y  -  t   -  r  g   i  s  t   r  i  s  g   t  c  s
 901  gacatgggct cttctccggt agcggcggag cttctacatc cgagccctgc
       g  h  g   l  f  s  g   s  g  g   a  s  t   s  e  p  c
 951  tcccatgcct ccagcgactc atggtcgctc ggcagctcct tgctcctaac
        s  h  a   s  s  d   s  w  s  l   g  s  s   l  l  l
1001  agtggaggcc agacttaggc acagcacgat gcccaccacc accagtgtgc
       t  v  e  a   r  l  r   h  s  t   m  p  t  t   t  s  v
1051  cgcacaaggc cgtggcggta gggtatgtgt ctgaaaatga gctcggggag
       p  h  k   a  v  a  v   g  y  v   s  e  n   e  l  g  e
1101  cgggcttgca ccgctgacgc atttggaaga cttaaggcag cggcagaaga
```

Figure 13 (Cont'd)

```
      r  a  c   t  a  d   a  f  g  r   l  k  a   a  a  e
1151  agatgcaggc agctgagttg ttgtgttctg ataagagtca gaggtaactc
      e  d  a  g   s  -  v   v  v  f   -  -  e  s   e  v  t
1201  ccgttgcggt gctgttaacg gtggagggca gtgtagtctg agcagtactc
      p  v  a   v  l  l  t   v  e  g   s  v  v   -  a  v  l
1251  gttgctgccg cgcgcgccac cagacataat agctgacaga ctaacagact
      v  a  a   a  r  a   t  r  h  n   s  -  q   t  n  r
1301  gttcctttcc atgggacttt tctgcagtca ccgtccaagc ttggtaccgc
      l  f  l  s   m  g  l   f  c  s   h  r  p  s   l  v  p
1351  caccaatgtg ggccaacgga acgtcaacgg taggcaaggc attcgcgctc
      p  p  m   w  a  n  g   t  s  t   v  g  k   a  f  a  l
1401  tcggcattca ttgtcggcgg gctgtggatc atcaccaccc agcacgtgaa
      s  a  f   i  v  g   g  l  w  i   i  t  t   q  h  v
1451  caactacgca acggcgtggt cggcatacag cgcgaccccg tactcgttgc
      n  n  y  a   t  a  w   s  a  y   s  a  t  p   y  s  l
1501  ttggcgtcgc cgaaattggc aagggcacgt tctcgagtat ggcgctggcc
      l  g  v   a  e  i  g   k  g  t   f  s  s   m  a  l  a
1551  acgattgtcg cgttgtacgc attcaccggt ttcgaatcga tcgcgaacgc
      t  i  v   a  l  y   a  f  t  g   f  e  s   i  a  n
1601  cgccgaagaa atggacgcgc cggaccggaa cctgccgaga gctataccga
      a  a  e  e   m  d  a   p  d  r   n  l  p  r   a  i  p
1651  tcgcgatctt ctcggttggc gcgatctact tgctcaccct aacggtagcg
      i  a  i   f  s  v  g   a  i  y   l  l  t   l  t  v  a
1701  atgctgctcg gatcgaacaa gatcgccgcg tcggacgaca ccgtgaaact
      m  l  l   g  s  n   k  i  a  a   s  d  d   t  v  k
1751  ggccgcggcc atcggaaacg ctaccttccg aacgatcatc gtcgtcggag
      l  a  a  a   i  g  n   a  t  f   r  t  i  i   v  v  g
1801  ccctgatatc gatgttcggc atcaatgtcg cggcctcgtt cggtgcaccg
      a  l  i   s  m  f  g   i  n  v   a  a  s   f  g  a  p
1851  cggctttgga ccgcgttagc ggacggatcc gaattcatgc atgaggtggc
      r  l  w   t  a  l   a  d  g  s   e  f  m   h  e  v
1901  tgctcgtgag caacgttcgg acgggccgat gaggctggat gcgcagggcc
      a  a  r  e   q  r  s   d  g  p   m  r  l  d   a  q  g
1951  gactgcagcg ttacgaggag gcgttcgctg actacgatgc accgtttgcg
      r  l  q   r  y  e  e   a  f  a   d  y  d   a  p  f  a
2001  ttcgtagatc tcgacgcgat gtggggcaat gccgatcaac tgcttgcgcg
      f  v  d   l  d  a   m  w  g  n   a  d  q   l  l  a
2051  cgccggcgac aagccgatcc gggtggcgtc gaagtcgctg cgttgccgac
      r  a  g  d   k  p  i   r  v  a   s  k  s  l   r  c  r
2101  cactgcaacg cgaaatcctt gatgccagtg agcgattcga cgggctattg
      p  l  q   r  e  i  l   d  a  s   e  r  f   d  g  l  l
2151  acgttcacgc ttaccgagac gctgtggctt gccggccaag gtttctcgaa
      t  f  t   l  t  e   t  l  w  l   a  g  q   g  f  s
2201  cctgttgttg gcctacccgc cgaccgaccg ggcggcattg cgtgcgcttg
      n  l  l  l   a  y  p   p  t  d   r  a  a  l   r  a  l
2251  gcgagctgac ggccaaggac ccggacgggg cgccgatcgt gatggtggac
      g  e  l   t  a  k  d   p  d  g   a  p  i   v  m  v  d
2301  agcgtggagc accttgacct gatcgagcgc acgaccgaca agccggtacg
      s  v  e   h  l  d   l  i  e  r   t  t  d   k  p  v
2351  gctgtgtctg gatttcgatg ccggctattg gcgcgccggc gggcggataa
      r  l  c  l   d  f  d   a  g  y   w  r  a  g   g  r  i
```

Figure 13 (Cont'd)

```
2401  aatgatctag agggccctat tctatagtgt cacctaaatg ctagagctcg
       k  -  s   r  g  p  y   s  i  v   s  p  k   c  -  s  s
2451  ctgatcagcc tcgactgtgc cttctagttg ccagccatct gttgtttgcc
        l  i  s   l  d  c   a  f  -  l   p  a  i   c  c  l
2501  cctcccccgt gccttccttg accctggaag gtgccactcc cactgtcctt
      p  l  p  r   a  f  l   d  p  g   r  c  h  s   h  c  p
2551  tcctaataaa atgaggaaat tgcatcgcat tgtctgagta ggtgtcattc
        f  l  i   k  -  g  n   c  i  a   l  s  e   -  v  s  f
2601  tattctgggg ggtggggtgg ggcaggacag caagggggag gattgggaag
        y  s  g   g  w  g   g  a  g  q   q  g  g   g  l  g
2651  acaatagcag gcatgctggg gatgcggtgg gctctatggc ttctgaggcg
      r  q  -  q   a  c  w   g  c  g   g  l  y  g   f  -  g
2701  gaaagaacca gctggggctc tagggggtat ccccacgcgc cctgtagcgg
        g  k  n   q  l  g  l   -  g  v   s  p  r   a  l  -  r
2751  cgcattaagc gcggcgggtg tggtggttac gcgcagcgtg accgctacac
        r  i  k   r  g  g   c  g  g  y   a  q  r   d  r  y
2801  ttgccagcgc cctagcgccc gctcctttcg ctttcttccc ttcctttctc
      t  c  q  r   p  s  a   r  s  f   r  f  l  p   f  l  s
2851  gccacgttcg ccggctttcc ccgtcaagct ctaaatcggg gcatcccttt
        r  h  v   r  r  l  s   p  s  s   s  k  s   g  h  p  f
2901  agggttccga tttagtgctt tacggcacct cgaccccaaa aaacttgatt
        r  v  p   i  -  c   f  t  a  p   r  p  q   k  t  -
2951  agggtgatgg ttcacgtagt gggccatcgc cctgatagac ggtttttcgc
      l  g  -  w   f  t  -   w  a  i   a  l  i  d   g  f  s
3001  cctttgacgt tggagtccac gttctttaat agtggactct tgttccaaac
      p  f  d   v  g  v  h   v  l  -   -  w  t   l  v  p  n
3051  tggaacaaca ctcaacccta tctcggtcta ttcttttgat ttataaggga
        w  n  n   t  q  p   y  l  g  l   f  f  -   f  i  r
3101  ttttggggat ttcggcctat tggttaaaaa atgagctgat ttaacaaaaa
      d  f  g  d   f  g  l   l  v  k   k  -  a  d   l  t  k
3151  tttaacgcga attaattctg tggaatgtgt gtcagttagg gtgtggaaag
        i  -  r   e  l  i  l   w  n  v   c  q  l   g  c  g  k
3201  tccccaggct ccccaggcag gcagaagtat gcaaagcatg catctcaatt
        s  p  g   s  p  g   r  q  k  y   a  k  h   a  s  q
3251  agtcagcaac caggtgtgga aagtccccag gctccccagc aggcagaagt
      l  v  s  n   q  v  w   k  v  p   r  l  p  s   r  q  k
3301  atgcaaagca tgcatctcaa ttagtcagca accatagtcc cgcccctaac
        y  a  k   h  a  s  q   l  v  s   n  h  s   p  a  p  n
3351  tccgcccatc ccgcccctaa ctccgcccag ttccgcccat tctccgcccc
        s  a  h   p  a  p   n  s  a  q   f  r  p   f  s  a
3401  atggctgact aatttttttt atttatgcag aggccgaggc cgcctctgcc
      p  w  l  t   n  f  f   y  l  c   r  g  r  g   r  l  c
3451  tctgagctat tccagaagta gtgaggaggc ttttttggag gcctaggctt
        l  -  a   i  p  e  v   v  r  r   l  f  w   r  p  r  l
3501  ttgcaaaaag ctcccgggag cttgtatatc cattttcgga tctgatcaag
        l  q  k   a  p  g   s  l  y  i   h  f  r   i  -  s
3551  agacaggatg aggatcgttt cgcatgattg aacaagatgg attgcacgca
      r  d  r  m   r  i  v   s  h  d   -  t  r  w   i  a  r
3601  ggttctccgg ccgcttgggt ggagaggcta ttcggctatg actgggcaca
        r  f  s   g  r  l  g   g  e  a   i  r  l   -  l  g  t
3651  acagacaatc ggctgctctg atgccgccgt gttccggctg tcagcgcagg
```

Figure 13 (Cont'd)

```
          t  d  n   r  l  l   -  c  r  r   v  p  a   v  s  a
3701    ggcgcccggt tctttttgtc aagaccgacc tgtccggtgc cctgaatgaa
          g  a  p  g   s  f  c   q  d  r   p  v  r  c   p  e  -
3751    ctgcaggacg aggcagcgcg gctatcgtgg ctggccacga cgggcgttcc
          t  a  g   r  g  s  a   a  i  v   a  g  h   d  g  r  s
3801    ttgcgcagct gtgctcgacg ttgtcactga agcgggaagg gactggctgc
          l  r  s   c  a  r   r  c  h  -   s  g  k   g  l  a
3851    tattgggcga agtgccgggg caggatctcc tgtcatctca ccttgctcct
          a  i  g  r   s  a  g   a  g  s   p  v  i  s   p  c  s
3901    gccgagaaag tatccatcat ggctgatgca atgcggcggc tgcatacgct
          c  r  e   s  i  h  h   g  -  c   n  a  a   a  a  y  a
3951    tgatccggct acctgcccat tcgaccacca agcgaaacat cgcatcgagc
          -  s  g   y  l  p   i  r  p  p   s  e  t   s  h  r
4001    gagcacgtac tcggatggaa gccggtcttg tcgatcagga tgatctggac
          a  s  t  y   s  d  g   s  r  s   c  r  s  g   -  s  g
4051    gaagagcatc aggggctcgc gccagccgaa ctgttcgcca ggctcaaggc
          r  r  a   s  g  a  r   a  s  r   t  v  r   q  a  q  g
4101    gcgcatgccc gacggcgagg atctcgtcgt gacccatggc gatgcctgct
          a  h  a   r  r  r   g  s  r  r   d  p  w   r  c  l
4151    tgccgaatat catggtggaa aatggccgct tttctggatt catcgactgt
          l  a  e  y   h  g  g   k  w  p   l  f  w  i   h  r  l
4201    ggccggctgg gtgtggcgga ccgctatcag gacatagcgt tggctacccg
          w  p  a   g  c  g  g   p  l  s   g  h  s   v  g  y  p
4251    tgatattgct gaagagcttg gcggcgaatg ggctgaccgc ttcctcgtgc
          -  y  c   -  r  a   w  r  r  m   g  -  p   l  p  r
4301    tttacggtat cgccgctccc gattcgcagc gcatcgcctt ctatcgcctt
          a  l  r  y   r  r  s   r  f  a   a  h  r  l   l  s  p
4351    cttgacgagt tcttctgagc gggactctgg ggttcgaaat gaccgaccaa
          s  -  r   v  l  l  s   g  t  l   g  f  e   m  t  d  q
4401    gcgacgccca acctgccatc acgagatttc gattccaccg ccgccttcta
          a  t  p   n  l  p   s  r  d  f   d  s  t   a  a  f
4451    tgaaaggttg ggcttcggaa tcgttttccg ggacgccggc tggatgatcc
          y  e  r  l   g  f  g   i  v  f   r  d  a  g   w  m  i
4501    tccagcgcgg ggatctcatg ctggagttct tcgcccaccc caacttgttt
          l  q  r   g  d  l  m   l  e  f   f  a  h   p  n  l  f
4551    attgcagctt ataatggtta caaataaagc aatagcatca caaatttcac
          i  a  a   y  n  g   y  k  -  s   n  s  i   t  n  f
4601    aaataaagca tttttttcac tgcattctag ttgtggtttg tccaaactca
          t  n  k  a   f  f  s   l  h  s   s  c  g  l   s  k  l
4651    tcaatgtatc ttatcatgtc tgtataccgt cgacctctag ctagagcttg
          i  n  v   s  y  h  v   c  i  p   s  t  s   s  -  s  l
4701    gcgtaatcat ggtcatagct gtttcctgtg tgaaattgtt atccgctcac
          a  -  s   w  s  -   l  f  p  v   -  n  c   y  p  l
4751    aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctggggtg
          t  i  p  h   n  i  r   a  g  s   i  k  c  k   a  w  g
4801    cctaatgagt gagctaactc acattaattg cgttgcgctc actgcccgct
          a  -  -   v  s  -  l   t  l  i   a  l  r   s  l  p  a
4851    ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg
          f  q  s   g  n  l   s  c  q  l   h  -  -   i  g  q
4901    cgcggggaga ggcggtttgc gtattgggcg ctcttccgct tcctcgctca
          r  a  g  r   g  g  l   r  i  g   r  s  s  a   s  s  l
4951    ctgactcgct gcgctcggtc gttcggctgc ggcgagcggt atcagctcac
```

Figure 13 (Cont'd)

```
      t  d  s   l  r  s  v   v  r  l   r  r  a   v  s  a  h
5001 tcaaaggcgg taatacggtt atccacagaa tcaggggata acgcaggaaa
      s  k  a   v  i  r   l  s  t  e   s  g  d   n  a  g
5051 gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg
     k  n  m  -   a  k  g   q  q  k   a  r  n  r   k  k  a
5101 cgttgctggc gtttttccat aggctccgcc cccctgacga gcatcacaaa
      a  l  l   a  f  f  h   r  l  r   p  p  d   e  h  h  k
5151 aatcgacgct caagtcagag gtggcgaaac ccgacaggac tataaagata
      n  r  r   s  s  q   r  w  r  n   p  t  g   l  -  r
5201 ccaggcgttt ccccctggaa gctccctcgt gcgctctcct gttccgaccc
     y  q  a  f   p  p  g   s  s  l   v  r  s  p   v  p  t
5251 tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg
      l  p  l   t  g  y  l   s  a  f   l  p  s   g  s  v  a
5301 ctttctcaat gctcacgctg taggtatctc agttcggtgt aggtcgttcg
      l  s  q   c  s  r   c  r  y  l   s  s  v   -  v  v
5351 ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg
     r  s  k  l   g  c  v   h  e  p   p  v  q  p   d  r  c
5401 ccttatccgg taactatcgt cttgagtcca acccggtaag acacgactta
      a  l  s   g  n  y  r   l  e  s   n  p  v   r  h  d  l
5451 tcgccactgg cagcagccac tggtaacagg attagcagag cgaggtatgt
      s  p  l   a  a  a   t  g  n  -   i  s  r   a  r  y
5501 aggcggtgct acagagttct tgaagtggtg gcctaactac ggctacacta
     v  g  g  a   t  e  f   l  k  w   w  p  n  y   g  y  t
5551 gaaggacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga
      r  r  t   v  f  g  i   c  a  l   l  k  p   v  t  f  g
5601 aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg
      k  r  v   g  s  s   -  s  g  k   q  t  t   a  g  s
5651 tggttttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc
     g  g  f  f   v  c  k   q  q  i   t  r  r  k   k  g  s
5701 aagaagatcc tttgatcttt tctacggggt ctgacgctca gtggaacgaa
      q  e  d   p  l  i  f   s  t  g   s  d  a   q  w  n  e
5751 aactcacgtt aagggatttt ggtcatgaga ttatcaaaaa ggatcttcac
      n  s  r   -  g  i   l  v  m  r   l  s  k   r  i  f
5801 ctagatcctt ttaaattaaa aatgaagttt taaatcaatc taaagtatat
     t  -  i  l   l  n  -   k  -  s   f  k  s  i   -  s  i
5851 atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct
     y  e  -   t  w  s  d   s  y  q   c  l  i   s  e  a  p
5901 atctcagcga tctgtctatt tcgttcatcc atagttgcct gactccccgt
      i  s  a   i  c  l   f  r  s  s   i  v  a   -  l  p
5951 cgtgtagata actacgatac gggagggctt accatctggc cccagtgctg
     v  v  -  i   t  t  i   r  e  g   l  p  s  g   p  s  a
6001 caatgatacc gcgagaccca cgctcaccgg ctccagattt atcagcaata
      a  m  i   p  r  d  p   r  s  p   a  p  d   l  s  a  i
6051 aaccagccag ccggaagggc cgagcgcaga agtggtcctg caactttatc
      n  q  p   a  g  r   a  e  r  r   s  g  p   a  t  l
6101 cgcctccatc cagtctatta attgttgccg ggaagctaga gtaagtagtt
     s  a  s  i   q  s  i   n  c  c   r  e  a  r   v  s  s
6151 cgccagttaa tagtttgcgc aacgttgttg ccattgctac aggcatcgtg
      s  p  v   n  s  l  r   n  v  v   a  i  a   t  g  i  v
6201 gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg
      v  s  r   s  s  f   g  m  a  s   f  s  s   g  s  q
6251 atcaaggcga gttacatgat cccccatgtt gtgcaaaaaa gcggttagct
```

Figure 13 (Cont'd)

```
      r  s  r  r   v  t  -   s  p  m   l  c  k  k   a  v  s
6301  ccttcggtcc tccgatcgtt gtcagaagta agttggccgc agtgttatca
       s  f  g   p  p  i  v   v  r  s   k  l  a   a  v  l  s
6351  ctcatggtta tggcagcact gcataattct cttactgtca tgccatccgt
       l  m  v   m  a  a   l  h  n  s   l  t  v   m  p  s
6401  aagatgcttt tctgtgactg gtgagtactc aaccaagtca ttctgagaat
      v  r  c  f   s  v  t   g  e  y   s  t  k  s   f  -  e
6451  agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat
       -  c  m   r  r  p  s   c  s  c   p  a  s   i  r  d  n
6501  accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc
       t  a  p   h  s  r   t  l  k  v   l  i  i   g  k  r
6551  ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga
      s  s  g  r   k  l  s   r  i  l   p  l  l  r   s  s  s
6601  tgtaacccac tcgtgcaccc aactgatctt cagcatcttt tactttcacc
       m  -  p   t  r  a  p   n  -  s   s  a  s   f  t  f  t
6651  agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg caaaaaaggg
       s  v  s   g  -  a   k  t  g  r   q  n  a   a  k  k
6701  aataagggcg acacggaaat gttgaatact catactcttc ctttttcaat
      g  i  r  a   t  r  k   c  -  i   l  i  l  f   l  f  q
6751  attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt
       y  y  -   s  i  y  q   g  y  c   l  m  s   g  y  i  f
6801  gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg
       e  c  i   -  k  n   k  q  i  g   v  p  r   t  f  p
6851  aaaagtgcca cctgacgtcg acggatcggg agatctcccg atcccctatg
      r  k  v  p   p  d  v   d  g  s   g  d  l  p   i  p  y
6901  gtcgactctc agtacaatct gctctgatgc cgcatagtta agccagtatc
       g  r  l   s  v  q  s   a  l  m   p  h  s   -  a  s  i
6951  tgctccctgc ttgtgtgttg gaggtcgctg agtagtgcgc gagcaaaatt
       c  s  l   l  v  c   w  r  s  l   s  s  a   r  a  k
7001  taagctacaa caaggcaagg cttgaccgac aattgcatga agaatctgct
      f  k  l  q   q  g  k   a  -  p   t  i  a  -   r  i  c
7051  tagggttagg cgttttgcgc tgcttcgcga tgtacgggcc agatatacgc
       l  g  l   g  v  l  r   c  f  a   m  y  g   p  d  i  r
7101  gttgacattg attattgact agttattaat agtaatcaat tacggggtca
       v  d  i   d  y  -   l  v  i  n   s  n  q   l  r  g
7151  ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa
      h  -  f  i   a  h  i   w  s  s   a  l  h  n   l  r  -
7201  tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa
       m  a  r   l  a  d  r   p  t  t   p  a  h   -  r  q  -
7251  tgacgtatgt tcccatagta acgccaatag ggactttcca ttgacgtcaa
       -  r  m   f  p  -   -  r  q  -   g  l  s   i  d  v
7301  tgggtggact atttacggta aactgcccac ttggcagtac atcaagtgta
      n  g  w  t   i  y  g   k  l  p   t  w  q  y   i  k  c
7351  tcatatgcca agtacgcccc ctattgacgt c          SEQ ID NO 17
       i  i  c   q  v  r  p   l  l  t            SEQ ID NO 18
```

Figure 14

First half of fusion insert from Open Reading Frame Rv1979c

27 Mar 2003 Feature Properties

Molecule: DNA Fusion Vaccine New, 7381 bps DNA
Circular

File Name: 1979-1769Fusion.cm5, dated 29 Jul 2002
Description: Ligation of inverted Rv1769 3' New* into
first step

Details: 'Rv1979c', Gene, 1362 to 1874
hypothetical protein Rv1979c
Translation product 171 aas
Mol Wt 17639.0, Isoelectric Pt (pI) 5.57

Translation:

WANGTSTVGKAFALSAFIVGGLWIITTQHVNNYATAWSAYSATPYSLLGVA

EIGKGTFSSMALATIVALYAFTGFESIANAAEEMDAPDRNLPRAIPIAIF

SVGAIYLLTLTVAMLLGSNKIAASDDTVKLAAAIGNATFRTIIVVGALIS

MFGINVAASFGAPRLWTALAD SEQ ID NO 19

Position within the ORF of the segment to be fused in the vaccine (BOLD)

M. tuberculis bacteria|Rv1979c|Rv1979c: 481 aa - POSSIBLE CONSERVED PERMEASE

1 - VVGPRTRGYA IHKLGFCSVV MLGINSIIGA GIFLTPGEVI GLAGPFAPMA YVLAGIFAGV
61 - VAIVFATAAR YVRTNGASYA YTTAAFGRRI GIYVGVTHAI TASIAWGVLA SFFVSTLLRV
121 - AFPDKAWADA EQLFSVKTLT FLGFIGVLLA INLFGNRAIK **WANGTSTVGK AFALSAFIVG**
181 - **GLWIITTQHV NNYATAWSAY SATPYSLLGV AEIGKGTFSS MALATIVALY AFTGFESIAN**
241 - **AAEEMDAPDR NLPRAIPIAI FSVGAIYLLT LTVAMLLGSN KIAASDDTVK LAAAIGNATF**
301 - **RTIIVVGALI SMFGINVAAS FGAPRLWTAL** ADSGVLPTRL SRKNQYDVPM VSFAITASLA
361 - LAFPLALRFD NLHLTGLAVI ARFVQFIIVP IALIALARSQ AVEHAAVRRN AFTDKVLPLV
421 - AIVVSVGLAV SYDYRCIFLV RGGPNYFSIA LIVITFVVVP AMAYLHYYRI IRRVGDRPST
481 - R SEQ ID NO 20

Figure 15

Second half of fusion insert from Open Reading Frame Rv1769

27 Mar 2003 Feature Properties

Molecule: DNA Fusion Vaccine New, 7381 bps DNA
Circular

File Name: 1979-1769Fusion.cm5, dated 29 Jul 2002
Description: Ligation of inverted Rv1769 3' New* into
first step

Details: 'Rv1769', Gene, 1890 to 2402
hypothetical protein Rv1769
Translation product 171 aas
Mol Wt 19105.2, Isoelectric Pt (pI) 4.78

Translation:

VHEVAAREQRSDGPMRLDAQGRLQRYEEAFADYDAPFAFVDLDAMWGNADQ

LLARAGDKPIRVASKSLRCRPLQREILDASERFDGLLTFTLTETLWLAGQ

GFSNLLLAYPPTDRAALRALGELTAKDPDGAPIVMVDSVEHLDLIERTTD

KPVRLCLDFDAGYWRAGGRIK SEQ ID NO 21

Position within the ORF of the segment to be fused in the vaccine (BOLD)

>M. tuberculis bacteria|Rv1769|Rv1769: 414 aa - CONSERVED HYPOTHETICAL PROTEIN
SEQ ID NO 22

1 - **VHEVAAREQR SDGPMRLDAQ GRLQRYEEAF ADYDAPFAFV DLDAMWGNAD QLLARAGDKP**
61 - **IRVASKSLRC RPLQREILDA SERFDGLLTF TLTETLWLAG QGFSNLLLAY PPTDRAALRA**
121 - **LGELTAKDPD GAPIVMVDSV EHLDLIERTT DKPVRLCLDF DAGYWRAGGR** IKIGSKRSPL
181 - HTPEQARALA VEIARRPALT LAALMCYEAH IAGLGDNVAG KRVHNAIIRR MQRMSFEELR
241 - ERRARAVELV REVADIKIVN AGGTGDLQLV AQEPLITEAT AGSGFYAPTL FDSYSTFTLQ
301 - PAAMFALPVC RRPGAKTVTA LGGGYLASGV GAKDRMPTPY LPVGLKLNAL EGTGEVQTPL
361 - SGDAARRLKL GDKVYFRHTK AGELCERFDH LHLVRGAEVV DTVPTYRGEG RTFL

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 November 2003 (13.11.2003)

PCT

(10) International Publication Number
WO 2003/093307 A2

(51) International Patent Classification[7]: C07K 14/35, G01N 33/569, C12N 15/31, A61K 38/16, 39/04

(21) International Application Number: PCT/GB2003/001815

(22) International Filing Date: 28 April 2003 (28.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0209723.6 27 April 2002 (27.04.2002) GB
0209724.4 27 April 2002 (27.04.2002) GB

(71) Applicant (*for all designated States except US*): THE SECRETARY OF STATE FOR ENVIRONMENT, FOOD AND RURAL AFFAIRS [GB/GB]; (DEFRA), Nobel House, 17 Smith Square, London SW1P 3JR, acting through the Veterinary Laboratories, Agency of New Haw, Addlestone, Surrey KT15 3NB (GB).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **COCKLE, Paul, Jason** [GB/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB). **VORDERMEIER, Hanns, Martin** [DE/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB). **GORDON, Stephen, Vincent** [IE/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB). **HEWINSON, Robert, Glyn** [GB/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB).

(74) Agent: **GREAVES, Carol, Pauline**; Greaves Brewster, Indigo House, Cheddar Business Park, Wedmore Road, Cheddar, Somerset BS27 3EB (GB).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

(48) Date of publication of this corrected version: 12 February 2004

(15) Information about Correction:
see PCT Gazette No. 07/2004 of 12 February 2004, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2003/093307 A2

(54) Title: MYCOBACTERIAL ANTIGENS AND USES THEREOF

(57) Abstract: The present invention relates to the use of antigens derived from the RD1 or RD2 regions of the *Mycobacterium tuberculosis*, *Mycobacterium bovis* or *Mycobacterium africanum* genomes, and peptides derived therefrom, as diagnostic reagents, in particular in the context of diagnostic kits. In addition, certain of these peptides, as well as other antigens and peptides derived from the RD14 region of the genome are suitable for use as vaccines. Novel fusion peptides are also part of the invention.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 November 2003 (13.11.2003)

PCT

(10) International Publication Number
WO 2003/093307 A3

(51) International Patent Classification[7]: C07K 14/35, G01N 33/569, C12N 15/31, A61K 38/16, 39/04

(21) International Application Number: PCT/GB2003/001815

(22) International Filing Date: 28 April 2003 (28.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0209723.6 27 April 2002 (27.04.2002) GB
0209724.4 27 April 2002 (27.04.2002) GB

(71) Applicant *(for all designated States except US)*: **THE SECRETARY OF STATE FOR ENVIRONMENT, FOOD AND RURAL AFFAIRS** [GB/GB]; (DEFRA), Nobel House, 17 Smith Square, London SW1P 3JR, acting through the Veterinary Laboratories, Agency of New Haw, Addlestone, Surrey KT15 3NB (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **COCKLE, Paul, Jason** [GB/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB). **VORDERMEIER, Hanns, Martin** [DE/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB). **GORDON, Stephen, Vincent** [IE/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB). **HEWINSON, Robert, Glyn** [GB/GB]; Veterinary Laboratories Agency, New Haw, Addlestone, Surrey KT15 3NB (GB).

(74) Agent: **GREAVES, Carol, Pauline**; Greaves Brewster, Indigo House, Cheddar Business Park, Wedmore Road, Cheddar, Somerset BS27 3EB (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 18 March 2004

(15) Information about Correction:
Previous Correction:
see PCT Gazette No. 07/2004 of 12 February 2004, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2003/093307 A3

(54) Title: MYCOBACTERIAL ANTIGENS AND USES THEREOF

(57) Abstract: The present invention relates to the use of antigens derived from the RD1 or RD2 regions of the *Mycobacterium tuberculosis*, *Mycobacterium bovis* or *Mycobacterium africanum* genomes, and peptides derived therefrom, as diagnostic reagents, in particular in the context of diagnostic kits. In addition, certain of these peptides, as well as other antigens and peptides derived from the RD14 region of the genome are suitable for use as vaccines. Novel fusion peptides are also part of the invention.

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB 03/01815

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/35 G01N33/569 C12N15/31 A61K38/16 A61K39/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, EMBASE, BIOSIS, MEDLINE, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	VAN PINXTEREN ET AL.,: "Diagnosis of Tuberculosis Based on the Two Specific Antigens ESAT-6 and CFP10" CLINICAL AND DIAGNOSTIC LABORATORY IMMUNOLOGY, vol. 7, no. 2, March 2000 (2000-03), pages 155-160, XP002252151 cited in the application the whole document -----	1,2, 13-19
X	WO 00 66157 A (GENNARO MARIA L;PUBLIC HEALTH RES INST OF THE) 9 November 2000 (2000-11-09) the whole document in particular pages 11-13 -----	1,2, 13-21, 26-28, 31-33

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 September 2003	13.11.03

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Roscoe, R.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 03/01815

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00 11214 A (BEHR MARCEL ;UNIV LELAND STANFORD JUNIOR (US); SMALL PETER (US); S) 2 March 2000 (2000-03-02) the whole document in particular pages 16-19	1,2, 13-19
X,P	COCKLE P J ET AL: "Identification of novel Mycobacterium tuberculosis antigens with potential as diagnostic reagents or subunit vaccine candidates by comparative genomics." INFECTION AND IMMUNITY. UNITED STATES DEC 2002, vol. 70, no. 12, December 2002 (2002-12), pages 6996-7003, XP002252152 ISSN: 0019-9567 the whole document	1,2, 13-21, 26-28, 31-33
A	ANDERSON P: "TB Vaccines: Progress and Problems" TRENDS IN IMMUNOLOGY, vol. 22, no. 3, March 2001 (2001-03), pages 160-168, XP002252153 cited in the application the whole document	1,2, 13-21, 26-28, 31-33

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 03/01815

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claim(s) 17, 18 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Although claims 32, 33 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.5), should the problems which led to the Article 17(2) declaration be overcome.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1, 2, 13-21, 26-28, 31-33 (all part)

   Rv1986 protein and nucleic acids (Seq.ID Nos. 1, 8, 61). diagnostic reagents / kits comprising said protein(s), nucleic acids encoding said diagniostic reagents, methods for determining M. tuberculosis infection, Vaccines comprising the polypeptides or therefore encoding nucleic acids and vaccination methods with either.

   ---

2. claims: 7, 8 and 1-3, 13-19, 26-28, 31-33 (all part)

   Rv3878 protein and nucleic acids (Seq.ID Nos. 3, 10, 63 (& fragments 35-47)). diagnostic reagents / kits comprising said protein(s), nucleic acids encoding said diagniostic reagents, methods for determining M. tuberculosis infection, Vaccines comprising the polypeptides or therefore encoding nucleic acids and vaccination methods with either.

   ---

3. claims: 1, 2, 13-21, 26-28, 31-33 (all part)

   Rv1983 protein and nucleic acids (Seq.ID Nos. 4, 11, 64). diagnostic reagents / kits comprising said protein(s), nucleic acids encoding said diagniostic reagents, methods for determining M. tuberculosis infection, Vaccines comprising the polypeptides or therefore encoding nucleic acids and vaccination methods with either.

   ---

4. claims: 1-3, 13-19, 26-28, 31-33 (all part)

   Rv3873 protein and nucleic acids (Seq.ID Nos. 5, 12, 65 (& fragments 7, 23-34)). diagnostic reagents / kits comprising said protein(s), nucleic acids encoding said diagniostic reagents, methods for determining M. tuberculosis infection, Vaccines comprising the polypeptides or therefore encoding nucleic acids and vaccination methods with either.

   ---

5. claims: 4-6, 9-12 and 1-3, 13-19, 26-28, 31-33 (all part)

   Rv3879c protein and nucleic acids (Seq.ID Nos. 6, 13, 66 (& fragments 48-58)). diagnostic reagents / kits comprising said protein(s), nucleic acids encoding said diagniostic reagents, methods for determining M. tuberculosis infection, Vaccines comprising the polypeptides or therefore encoding nucleic acids and vaccination methods with either.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

6. claims: 19, 26-28, 31-33 (all part)

   Rv3872 protein and nucleic acids (Seq.ID Nos. 2, 9, 62). diagnostic reagents comprising said protein(s), nucleic acids encoding said diagniostic reagents, methods for determining M. tuberculosis infection, Vaccines comprising the polypeptides or therefore encoding nucleic acids and vaccination methods with either.

   ---

7. claims: 20-33 (all part)

   Rv1979c protein and nucleic acids (Seq.ID Nos. 14, 59, 70). diagnostic reagents comprising said protein(s), nucleic acids encoding said diagniostic reagents, methods for determining M. tuberculosis infection, Vaccines comprising the polypeptides or therefore encoding nucleic acids and vaccination methods with either. Further, fusion proteins of Rv1979c and Rv1769.

   ---

8. claims: 20-23, 26-33 (all part)

   Rv1769 protein and nucleic acids (Seq.ID Nos. 15, 60, 71). diagnostic reagents comprising said protein(s), nucleic acids encoding said diagniostic reagents, methods for determining M. tuberculosis infection, Vaccines comprising the polypeptides or therefore encoding nucleic acids and vaccination methods with either.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 03/01815

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0066157	A	09-11-2000	AU	4702300 A	17-11-2000
			CA	2372583 A1	09-11-2000
			EP	1214088 A1	19-06-2002
			JP	2003519467 T	24-06-2003
			WO	0066157 A1	09-11-2000
WO 0011214	A	02-03-2000	US	6291190 B1	18-09-2001
			AU	5394699 A	14-03-2000
			EP	1108060 A1	20-06-2001
			WO	0011214 A1	02-03-2000
			US	2002176873 A1	28-11-2002

Form PCT/ISA/210 (patent family annex) (July 1992)